

Scene-Dependency of Spatial Image Quality Metrics

Fry, E.

A PhD thesis awarded by the University of Westminster.

© Dr Edward Fry, 2020.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk

Scene-Dependency of Spatial Image Quality Metrics

Edward William Stewart Fry, B.Sc. (Hons)

School of Computer Science and Engineering

College of Design, Creative and Digital Industries

University of Westminster

A thesis submitted in partial fulfilment of the requirements of the University of
Westminster for the Degree of Doctor of Philosophy.

This research programme was completed within the Computational Vision and Imaging
Technology Research Group.

Abstract

This thesis is concerned with the measurement of spatial imaging performance and the modelling of spatial image quality in digital capturing systems. Spatial imaging performance and image quality relate to the objective and subjective reproduction of luminance contrast signals by the system, respectively; they are critical to overall perceived image quality.

The Modulation Transfer Function (MTF) and Noise Power Spectrum (NPS) describe the signal (contrast) transfer and noise characteristics of a system, respectively, with respect to spatial frequency. They are both, strictly speaking, only applicable to linear systems since they are founded upon linear system theory. Many contemporary capture systems use adaptive image signal processing, such as denoising and sharpening, to optimise output image quality. These non-linear processes change their behaviour according to characteristics of the input signal (i.e. the scene being captured). This behaviour renders system performance “scene-dependent” and difficult to measure accurately. The MTF and NPS are traditionally measured from test charts containing suitable predefined signals (e.g. edges, sinusoidal exposures, noise or uniform luminance patches). These signals trigger adaptive processes at uncharacteristic levels since they are unrepresentative of natural scene content. Thus, for systems using adaptive processes, the resultant MTFs and NPSs are not representative of performance “in the field” (i.e. capturing real scenes).

Spatial image quality metrics for capturing systems aim to predict the relationship between MTF and NPS measurements and subjective ratings of image quality. They cascade both measures with contrast sensitivity functions that describe human visual sensitivity with respect to spatial frequency. The most recent metrics designed for adaptive systems use MTFs measured using the dead leaves test chart that is more representative of natural scene content than the abovementioned test charts. This marks a step toward modelling image quality with respect to real scene signals.

This thesis presents novel scene-and-process-dependent MTFs (SPD-MTF) and NPSs (SPD-NPS). They are measured from imaged pictorial scene (or dead leaves target) signals to account for system scene-dependency. Further, a number of spatial image quality metrics are revised to account for capture system and visual scene-dependency. Their MTF and NPS parameters were substituted for SPD-MTFs and SPD-NPSs. Likewise, their standard visual functions were substituted for contextual detection (cCSF) or discrimination (cVPF) functions. In addition, two novel spatial image quality metrics are presented (the log Noise Equivalent Quanta (NEQ) and Visual log NEQ) that implement SPD-MTFs and SPD-NPSs.

The metrics, SPD-MTFs and SPD-NPSs were validated by analysing measurements from simulated image capture pipelines that applied either linear or adaptive image signal processing. The SPD-NPS measures displayed little evidence of measurement error, and the metrics performed most accurately when they used SPD-NPSs measured from images of scenes. The benefit of deriving SPD-MTFs from images of scenes was traded-off, however, against measurement bias. Most metrics performed most accurately with SPD-MTFs derived from dead leaves signals. Implementing the cCSF or cVPF did not increase metric accuracy.

The log NEQ and Visual log NEQ metrics proposed in this thesis were highly competitive, outperforming metrics of the same genre. They were also more consistent than the IEEE P1858 Camera Phone Image Quality (CPIQ) metric when their input parameters were modified. The advantages and limitations of all performance measures and metrics were discussed, as well as their practical implementation and relevant applications.

Contents

Abstract	i
Contents.....	iii
List of Figures	vi
List of Tables.....	xi
Acknowledgements	xii
Author's Declaration	xiii
Chapter 1 Introduction	1
1.1 Imaging Performance Measures and System Non-Linearity	5
1.2 Human Visual System (HVS) Modelling and Non-Linearity.....	8
1.3 Aims and Overview.....	9
1.4 Original Contributions to Knowledge	10
Chapter 2 Image Quality and System Performance	11
2.1 Defining Image Quality and Fidelity	11
2.2 Scene-Dependency	13
2.2.1 Imaging System Performance	14
2.2.2 Human Visual System.....	18
2.2.3 Observer Preference	20
2.3 Image Quality Attributes	21
2.3.1 Contrast and Tone	22
2.3.2 Sharpness and Resolution	25
2.3.3 Noise	36
2.4 Signal-to-Noise Measures	42
2.5 Visual Models	45
2.6 Summary	53
Chapter 3 Image Quality Metrics.....	54
3.1 Image Quality Metrics (IQM) Review	54
3.1.1 Signal Transfer Visual IQMs (STV-IQM)	58
3.1.2 Multivariate Formalism IQMs (MF-IQM)	61
3.1.3 Image Fidelity IQMs (IF-IQM).....	64
3.1.4 Computational IQMs (CP-IQM)	66
3.2 Proposed Image Quality Metric (IQM) Frameworks	69
3.2.1 Log Noise Equivalent Quanta (log NEQ), Visual log NEQ, and Scene-and- Process-Dependent NEQ (SPD-NEQ)	70

3.2.2	Revised STV-IQMs.....	72
3.2.3	Revised Camera Phone Image Quality (CPIQ) Standard Metric	73
3.3	Summary	74
Chapter 4	Validation of Scene-and-Process-Dependent NPSs (SPD-NPS)	75
4.1	Derivation of the SPD-NPS Measures	75
4.2	Validation Methodology	81
4.3	Results	86
4.4	Summary	93
Chapter 5	Validation of Scene-and-Process-Dependent MTFs (SPD-MTF)	94
5.1	Derivation of the SPD-MTF Measures	94
5.2	Results	100
5.3	Summary	107
Chapter 6	An Evaluation of Scene-and-Process-Dependent IQMs	109
6.1	Methodology	109
6.1.1	Test Image Dataset	110
6.1.2	Psychophysical Evaluation.....	112
6.1.3	Generation of the IQM Variants	115
6.2	Results	117
6.2.1	Subjective Quality Ratings.....	117
6.2.2	Benchmarking of IQM Variants.....	122
6.2.3	Further Analysis of Selected IQM Variants.....	126
6.3	Summary	134
Chapter 7	Discussion	135
7.1	Image Quality Metric Frameworks	136
7.2	Scene-and-Process-Dependent Noise Power Spectra (SPD-NPS).....	138
7.3	Scene-and-Process-Dependent Modulation Transfer Functions (SPD-MTF) ...	140
7.4	Scene-and-Process-Dependent Image Quality Metrics.....	142
7.5	Implementation and Application.....	145
7.5.1	Practical Implementation of SPD-MTF and SPD-NPS measures.....	145
7.5.2	Application in Objective Imaging Performance Metrics	148
7.5.3	Application in Computer Vision and Autonomous Vehicles.....	150
Chapter 8	Conclusions and Recommendations for Further Work	152
8.1	Conclusions	152
8.2	Further Work.....	154
Appendix A	Linear System Theory Requirements	156
Appendix B	Structural Similarity Index (SSIM) Definition	157
Appendix C	Test Scenes used in Chapters 4 & 5	159

Appendix D	Test Scenes used in Chapter 6.....	163
Appendix E	Generation of Ruler Images by Allen	165
Appendix F	Display Characterisation	167
Appendix G	Snellen Near Vision Test Card.....	173
Appendix H	Observer Instructions	174
Appendix I	IQM Input Parameters	177
Appendix J	Mean Absolute Errors of IQMs.....	184
Appendix K	Related Work	191
K.1	List of Publications	191
K.2	Presentations at Conferences and Symposia	192
K.3	Awards	193
Abbreviations		195
References		200

List of Figures

Figure 1.1 A typical imaging chain; adapted from [1, p. 345]. ISP refers to image signal processing.....	1
Figure 1.2 Luminance (a) and chrominance (b) components of the Flower Garden image (c) [4] in the 8-bit digital YCbCr colour space.	2
Figure 1.3 Diagram summarising the function of capture system image quality metrics (IQM).	3
Figure 1.4 Characteristics of system performance measurements, $F(s)$, derived from a range of pictorial scene (or test chart) input signals s : a) describes a linear system, b) describes a system applying non-linear image signal processing (ISP). $F(s)$ is the mean of $F(s_1)$ to $F(s_n)$	3
Figure 1.5 Subjective image quality averaged over observer vs the Square Root Integral with Noise (SQRI _n) [12] metric at varying levels of non-linear Joint Photographic Experts Group (JPEG) compression. The JPEG quality factor of each image varies from value 100 to where subjective quality drops below 10% acceptability. Adapted from [13]; original data from [10].....	5
Figure 1.6 Examples of the types of input signals from test charts traditionally used in Modulation Transfer Function (MTF) ((a) to (c)), and Noise Power Spectrum (NPS) measurements (d).	6
Figure 1.7 MTF measurements obtained from a simulated capture system after linear and non-linear processing, from the following test targets: i) Imatest™ ISO 12233 [23] Slanted Edge Spatial Frequency Response (E-SFR) test chart [24] using sfrmat3 [25] (blue curves). ii) Imatest™ ISO 12233 sinusoidal Siemens Star test chart [26] using Image Engineering™ analysis software [27] (grey curves). iii) Imatest™ Spilled Coins (dead leaves) test chart using Burns' [28], [29] implementation of the direct dead leaves MTF [19] (orange curves). Both simulation pipelines are described in Section 4.2. The following processes were simulated first: lens blur, Poisson noise at a signal-to-noise ratio of 5, mosaicing and demosaicing [30]. The linear pipeline then applied Gaussian blurring (denoising), and unsharp masking (sharpening). The non-linear pipeline applied Block Matching and 3D Filtering [31] (denoising) and the Guided Image Filter [32] (sharpening).	7
Figure 2.1 Objective and subjective factors that affect perceived image quality.	12
Figure 2.2 Image fidelity and quality evaluation processes with respect to imaging chain distortions [1, p. 351]. Adapted from Ford [9, p. 8], and further from [54, p. 34].....	13
Figure 2.3 Scene-and-process-dependent noise images obtained using ten replicate captures following the method of Section 4.1. All output images and replicates were generated using a simulated image capture pipeline after Poisson noise was added at a linear signal-to-noise ratio (SNR) of 10 and non-linear content-aware denoising [31] and sharpening [84] were applied. Noise images (d), (e) and (f) correspond to the following input information: (a) a uniform-tone patch, (b) the 'Students' image [7], and (c) the 'Architecture' image [7], respectively (Appendix C). The contrast of each noise image was increased to emphasise scene-dependency.....	16

Figure 2.4 Demonstration of local-content-dependency in terms of signal transfer, caused by content-aware denoising and sharpening. Normalised pixel values, y , are plotted vs horizontal, x , and vertical pixel coordinates, z , for: (a) a two-dimensional (2D) input signal consisting of a low-contrast high-frequency signal ($0 < x \leq 15$), high contrast edge ($15 < x \leq 18$), and uniform tone signal ($18 < x \leq 30$). (b) shows (a) after adding Gaussian noise. (c) shows (b) after denoising with the Fast Bilateral Filter (FBF) [93]. (d) shows (c) after sharpening with the Weighted Least Squares (WLS) [92] filter.	17
Figure 2.5 Demonstration of human visual scene-dependency: (a) non-busy portrait [102], (b) busy landscape image [102]. (c) and (d) show the same images with Gaussian noise added at identical levels. Visual masking suppresses noisiness and perceived quality loss due to noise in (d).....	19
Figure 2.6 Visual saliency maps predicted by the Graph-Based Visual Saliency algorithm [103]. Adapted from [41, p. 43].	20
Figure 2.7 Sinusoidal modulations in terms of: a) luminance, and b) red-green chrominance signals; adapted from [117].	22
Figure 2.8 Heynacher & Kober's [147] resolving power versus sharpness example: (a) has higher resolving power, (b) has higher contrast at mid-high frequencies and is perceived to be sharper than (a) at most viewing distances.....	26
Figure 2.9 Relationships between the various spread functions and transfer functions that describe system signal transfer. u and v are spatial frequencies with respect to the x and y dimensions of the Point Spread Function (PSF), respectively. ω is spatial frequency with respect to the x dimension of the Line Spread Function (LSF) or Edge Spread Function (ESF). Adapted from [1, p. 133].	31
Figure 2.10 Test charts used to derive measurements of capture system signal transfer: (a) Imatest™ ISO 12233 [23] Slanted Edge Spatial Frequency Response (E-SFR) test chart [24]. (b) Image Engineering™ TE265 Dead Leaves test chart [151]. (c) Imatest™ Siemens Star test chart [26]. (d) Imatest™ Log Frequency-Contrast test chart [152].	32
Figure 2.11 Simulations by Farrell & Wandell [167] of: (a) temporal photon noise, and (b) read noise, as well as fixed-pattern noise caused by (c) photoresponse non-uniformity (PRNU) and (d) dark signal non-uniformity (DSNU) in image capture systems.	38
Figure 2.12 Image Engineering™ TE264 test chart [130] for noise measurements according to ISO 15739 [168]. The chart contains 20 uniform-tone patches.	39
Figure 2.13 Noise with identical mean signal and standard deviation (i.e. the same level of root mean square (RMS) noise) but varying correlation [41, p. 32]: (a) low two-dimensional (2D) covariance, (b) high 2D covariance, (c) low vertical covariance and high horizontal covariance.	39
Figure 2.14 Barten's detection (broken line) [186] and discrimination (solid line) [183, p. 143] contrast sensitivity functions (CSF) [187].	47
Figure 2.15 Isolated contrast detection (iCSF) [35] and contextual contrast detection (cCSF) [35] and discrimination (cVPF) functions [8] for the 'Buildings' (a) and 'Bench' (b) images shown in Appendix C. Adapted from [14].	50
Figure 3.1 Generalisation of Full-Reference and Reduced-Reference image quality metric (IQM) characteristics.	56
Figure 3.2 Generalisation of No-Reference (NR) image quality metric (IQM) characteristics.	56

Figure 3.3 Approximate relative levels of calibration employed by image quality metrics (IQM) from different genres and the Noise Equivalent Quanta (NEQ) system performance measure.	57
Figure 3.4 Diagram describing processing by the IEEE P1858 Camera Phone Image Quality (CPIQ) Standard's [22] implementation of Keelan's multivariate formalism IQM (MF-IQM) [46]. Input parameters are shown for the texture blur and visual noise attribute metrics only.	62
Figure 3.5 Generalisation of processing by threshold image fidelity metrics (IF-IQM), adapted from [264].	64
Figure 3.6 Generalisation of processing by suprathreshold image fidelity metrics (IF-IQM).	65
Figure 3.7 Diagram of processing by the Structural Similarity (SSIM) Index [114].	67
Figure 3.8 Diagram describing processing by the scene-and-process-dependent Noise Equivalent Quanta (SPD-NEQ) system performance measure.	71
Figure 3.9 Diagram describing processing by both the revised SQRIn [12] and PIC [169] metrics.	73
Figure 3.10 Diagram describing processing by the revised CPIQ metric [22] (with respect to texture blur and visual noise attribute metrics only).	74
Figure 4.1 The scene-and-process-dependent Noise Power Spectrum (SPD-NPS) framework.	76
Figure 4.2 Fixed patterns of artefacts caused by demosaicing the 'Students' image [7] (left), and the detail of it (right). Global contrast was enhanced to increase artefact intensity.	80
Figure 4.3 The linear (a) and non-linear (b) camera simulation pipeline modelling and image signal processing (ISP) stages. Output images were generated at points indicated by the red arrows. Modelling that was common to both pipelines is shaded grey. ISP stages for which the linear and non-linear algorithms had very different effects on image quality to one another are shaded blue.	81
Figure 4.4 Two-dimensional (2D) discrete Fourier transform (DFT) log luminance spectra for the 'Students' image [7]: before, (c), and after windowing, (d) using the mask (e).	86
Figure 4.5 Luminance Noise Power Spectra (NPS) and scene-and-process-dependent NPSs (SPD-NPS) derived from different test charts. The uniform patch NPS (grey curves) and proposed dead leaves SPD-NPS (red curves), are shown at different image signal processing (ISP) stages at SNR 40, (a) to (f), and SNR 5, (g) to (l). The power (y) axis is logarithmically scaled.	88
Figure 4.6 Scene-and-process-dependent luminance Noise Power Spectra (SPD-NPS) derived from pictorial images and the dead leaves test chart. The pictorial image SPD-NPS (grey curves), mean pictorial image SPD-NPS (black curves), pictorial image SPD-NPS standard deviation (black dotted curves), and dead leaves SPD-NPS (red curves) are shown after different stages of image signal processing (ISP) at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). The power (y) axis is logarithmically scaled.	89
Figure 4.7 Demonstration of luminance noise scene-dependency in the non-linear image capture pipeline. Pictorial image scene-and-process-dependent Noise Power Spectra (SPD-NPS) are plotted on linear axes for the linear and non-linear pipelines at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). Each curve is coloured according to its	

integrated area before denoising and sharpening. Green curves are of a higher area between zero and Nyquist frequency.	90
Figure 5.1 Scene-derived texture Modulation Transfer Functions (MTF) of Branca <i>et al.</i> [7] for the non-linear pipeline at a signal-to-noise ratio (SNR) of 40, before, (a) to (c), and after the windowing method of Section 4.2 was applied, (d) to (f). Test images are from Branca <i>et al.</i> [7], resized to 512 x 512 pixels by bicubic interpolation. Dotted curves show +/- 1 standard deviation.	97
Figure 5.2 The scene-and-process-dependent Modulation Transfer Function (SPD-MTF) framework.	98
Figure 5.3 Direct dead leaves Modulation Transfer Functions (MTF) (black curves) and proposed dead leaves scene-and-process-dependent MTFs (SPD-MTF) (red curves) after different stages of image signal processing (ISP) at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l).	101
Figure 5.4 Comparison of scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) from pictorial images and the dead leaves test chart. Pictorial image SPD-MTFs (grey curves), mean pictorial image SPD-MTFs (black curves), pictorial image SPD-MTF standard deviations (black dotted curves) and dead leaves SPD-MTFs (red curves) are shown after different image signal processing (ISP) stages at signal-to-noise ratios (SNR) 40, (a) to (f), and 5, (g) to (l).	102
Figure 5.5 Demonstration of signal transfer scene-dependency in the non-linear image capture pipeline. Pictorial image scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) are shown for the linear and non-linear pipelines at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). Curves for each scene image are coloured according to their integrated area before denoising and sharpening. Green curves are of a higher area between zero and Nyquist frequency. Ten replicates were used when computing each SPD-MTF.	103
Figure 6.1 Power spectra for the input images to the simulations (shown in Appendix D).	111
Figure 6.2 Layout of the image quality ruler graphics user interface (GUI) [102, Sec. 6.5.2]: (a) ruler image, (b) test image, (c) slider to select ruler images, (d) button to select next test image.	113
Figure 6.3 Layout of laboratory equipment (plan view).	114
Figure 6.4 Modulation Transfer Functions (MTF) of imaging chain components and Barten's optical MTF of the eye [183, p. 29]. The camera-lens, camera-lens-display and display MTFs were all measured by Allen [102, p. 212]. The latter was also modelled by Equation 3.2 [22, p. 16].	114
Figure 6.5 Mean observer image quality ratings on the Standard Quality Scale (SQS ₂) for each test image; error bars show standard error.	118
Figure 6.6 The scenes' susceptibilities to perceived quality loss. The x-axis is the grand mean of the scenes' quality ratings. The y-axis is the quality rating for each scene, averaged over all observers. Higher susceptibility scenes are distributed closer to the x-axis. The scenes classified as high, medium and low susceptibility are labelled [H], [M] and [L], respectively.	119
Figure 6.7 The observers' sensitivities to perceived quality loss. The x-axis is the grand mean of the observers' quality ratings. The y-axis is the quality rating for each observer, averaged over all scenes. Data from higher sensitivity observers is distributed closer to the	

x-axis. Observers classified as high, medium and low sensitivity are labelled [H], [M] and [L], respectively. Experienced and inexperienced observers are labelled [Y] and [N], respectively.	120
Figure 6.8 Box and whisker plots of Mean Absolute Errors (MAE) of all variants of each metric.....	123
Figure 6.9 Box and whisker plots of the Spearman's Rank Order Correlation Coefficient (SROCC) of all variants of each metric, including the Mean Structural Similarity (MSSIM).	123
Figure 6.10 The mean of all observer image quality ratings vs the log Noise Equivalent Quanta (NEQ) metric variants that were less accurate or sensitive to system scene-dependency. They implement direct dead leaves Modulation Transfer Functions (MTF) [19].	127
Figure 6.11 The mean of all observer image quality ratings vs log Noise Equivalent Quanta (log NEQ) variants that were more accurate and sensitive to system scene-dependency. They employ the pictorial image SPD-NPS. The most accurate log NEQ variant is shown in (a) and (b).	127
Figure 6.12 The mean of all observer image quality ratings vs the most accurate Visual log NEQ variants that both implemented the dead leaves SPD-MTF and mean pictorial image SPD-NPS.	128
Figure 6.13 The mean of all observer quality ratings vs the Perceived Information Capacity (PIC) [169] variants with lowest, (a) and (b), and highest accuracy, (c) and (d). Implementing different visual models had minimal effect on the correlations; hence the employed model is not specified.	128
Figure 6.14 The mean of all observer image quality ratings vs the Square Root Integral with noise (SQRIn) [12] variants of lowest, (a) and (b), and highest accuracy, (c) and (d). The visual model is not specified because implementing different models had minimal effect on the correlations.	129
Figure 6.15 The mean of all observer quality ratings vs CPIQ metric [22] variants that employed different contrast sensitivity functions (CSF). All variants implemented the direct dead leaves Modulation Transfer Function (MTF) [19] and uniform patch noise images that are default input parameters to the IEEE P1858 standard [22] and are least sensitive to system scene-dependency.	130
Figure 6.16 The mean of all observer image quality ratings vs CPIQ metric [22] variants that employed different Modulation Transfer Functions (MTF) and noise parameters. All variants implemented the Barten contrast sensitivity function (CSF) [186] that was most optimal for the CPIQ metric. The most accurate CPIQ metric variant is shown in (e) and (f).	131
Figure 6.17 The mean of all observer quality ratings vs output scores from the Mean Structural Similarity (MSSIM) metric [114].	132
Figure 7.1 Proposed automated display-capture device layout.	147
Figure 7.2 Mean pictorial image SPD-NPS area, in units of pixels.	149
Figure 7.3 Relative Standard Deviation Area of the SPD-NPSs of 50 input scenes. This is expressed as a percentage of the integrated area under the mean pictorial image SPD-NPS.	149

List of Tables

Table 2.1 Input parameters for Johnson and Fairchild’s luminance and chrominance CSFs [188] as implemented in the IEEE P1858 Camera Phone Image Quality (CPIQ) standard [22, p. 72].	48
Table 2.2 Summary of the most relevant performance measures concerning spatial image quality.	53
Table 3.1 Summary of image quality metric (IQM) genres; JND refers to just-noticeable difference.	55
Table 4.1 Summary of the uniform patch Noise Power Spectrum (NPS) and the scene-and-process-dependent NPS (SPD-NPS) measures. N is the number of replicate captures of each scene/chart.	77
Table 5.1 Summary of the direct dead leaves Modulation Transfer Function (MTF) measurement implementation [19] and the scene-and-process-dependent MTFs (SPD-MTF) of this thesis. PS and NPS are the power spectrum and Noise Power Spectrum, respectively. SPD-NPS is the scene-and-process-dependent NPS. M is the number of images in the test image set.	98
Table 5.2 R^2 correlation coefficients of a logarithmic curve fit of form $y = m \cdot \ln x + c$, to the regression between the integrated area under the pictorial images’ scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) and the busyness’ of these scenes [59].	107
Table 6.1 Summary of the contrast sensitivity functions (CSF) employed in Chapter 6.	109
Table 6.2 Optimal opacities for the pipelines’ Image Signal Processing (ISP) filters described in Section 4.2. BM3D, GIF, USM and Gaussian refer to the Block Matching with 3D Filtering [31], Guided Image Filter [32], MATLAB TM <code>imsharpen</code> unsharp mask [85] and Gaussian blur filters, respectively.	111
Table 6.3 Comparison of the conditions of the psychophysical image quality evaluations with the recommendations of ISO 20462 [56, Sec. 6.1].	115
Table 6.4 Input parameters and Mean Absolute Errors (MAE) for the most accurate (left table) and least accurate (right table) variants of each image quality metric (IQM). The MAE scores are coloured from red to green denoting lowest to highest accuracy.	123
Table 6.5 Input parameters and Spearman’s Rank Order Correlation Coefficients (SROCC) for variants of each image quality metric (IQM) that resulted in the highest (left table) and lowest correlation (right table), as well as the Mean Structural Similarity (MSSIM) metric. The SROCC scores are coloured from red to green denoting lowest to highest correlation, respectively.	124

Acknowledgements

This thesis marks the culmination of the efforts of several people who deserve due praise and recognition, not just the author.

Firstly, I thank my Director of Studies and mentor, Dr Sophie Triantaphillidou, for consistently showing belief in my development since my early undergraduate years. Her guidance and knowledge have been instrumental to this research project from the day it was conceptualised. I am also particularly grateful to the three further supervisors of this project, Professors Robin Jenkin, Ralph Jacobson and John Jarvis, for sharing the enormous depth and breadth of their understanding. Their ability as a team to provide a “helicopter view” over this multi-disciplinary project has been invaluable. Moreover, I thank the whole supervisory team for their kind support and encouragement.

I am also grateful to Dr Elizabeth Allen, Dr Peter Burns, and Riccardo Branca for providing their code for the ISO 20462 Image Quality Ruler interface, the direct dead leaves Modulation Transfer Function (MTF) implementation, and the scene-derived texture MTF implementation, respectively.

I must also thank my fellow PhD researchers, Oliver van Zwanenberg and Dr Vicent Sanchis-Jurado, for frequent, engaging conversations, as well as colleagues and friends at the University of Westminster and elsewhere, including Janko Smejkal, Dr Aleka Psarrou, Professor Tamas Kiss, John Smith, Dr Efthemia Bilissi, Dr Olivier Moullard, Dr Jae Young Park, Dr Anastasia Tsifouti and Dr Gaurav Gupta.

Finally, but by no means least, I am grateful to my brother George Fry, sister-in-law Claire Fry, and my girlfriend Jowita Nesterowicz for their unfailing support.

Dedicated to my parents

Jillie and Bernard

Author's Declaration

I, the undersigned, hereby declare that all the material contained in this thesis is my own work, expressed in my own words. It has not been submitted for any other degree or professional qualification. All sources of information, visual material, or data produced by other researchers has been referenced suitably, whether it has been used directly or indirectly.

Edward William Stewart Fry, B.Sc. (Hons)

February 2020

Chapter 1 Introduction

Evaluating the objective performance of capturing systems and their subjective image quality is necessary to their design, optimisation and comparison. Digital capturing systems record information about an input scene (or subject) as digital image files. Their optics form a continuous, but distorted image of the scene, from which information is extracted by the sensor and electronics discretely. Image signal processing (ISP) is often then applied before the digital image is encoded into a suitable storage format. To view this encoded image, it must be presented by an output system such as a display or printer. Each component in the resultant imaging chain contributes to the characteristics of the output image (Figure 1.1).

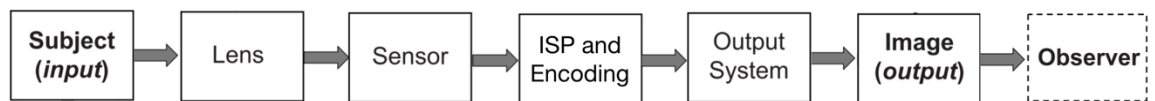


Figure 1.1 A typical imaging chain; adapted from [1, p. 345]. ISP refers to image signal processing.

Quantifying the relationship between the input scene and output image is a common, general aim when evaluating imaging system image quality. However, the input scene information is usually unavailable for capture systems, meaning output image quality must be quantified without reference to it.

Subjective (or perceived) image quality refers to the visual impression of goodness that an image conveys [1, p. 345]. It is broken down into a number of perceptual attributes (or dimensions) that are defined in Chapter 2. In this thesis, the term image quality refers specifically to the subjective (visual) image quality unless stated otherwise.

Objective image quality refers to physical measurements of imaging system performance, images or image attributes. These measurements provide a practical (and conceptual) means to assess and predict the effect of changes to the system [2, p. 1]. Certain measurements can be used to characterise systems, components, or processes with respect to a particular attribute, expressing their input-to-output characteristics as mathematical functions, or single figure variables. These measurements describe the physical properties of the

system/component/process that drive its objective performance, and subjective quality, with respect to the attribute in question.

The most useful performance measures correlate with overall perceived image quality or attribute strength. These relationships are well described in the prior art and have resulted in a number of standardised measurement methods discussed in Section 2.3. However, individual performance measures do not predict subjective image quality or attribute strength particularly well because the human visual system (HVS) is unaccounted for.

This thesis focuses on spatial image quality. This relates to the distribution of luminance contrast signals across the image, which are core to the perception of form, shape and detail, as shown in Figure 1.2(a). Spatial image quality is associated with the attributes of resolution, sharpness and noisiness that are primary drivers for the overall subjective image quality of contemporary capture systems [3]. The reproduction of chrominance signals (Figure 1.2(b)) also contributes significantly to overall subjective image quality but is not discussed in this thesis.

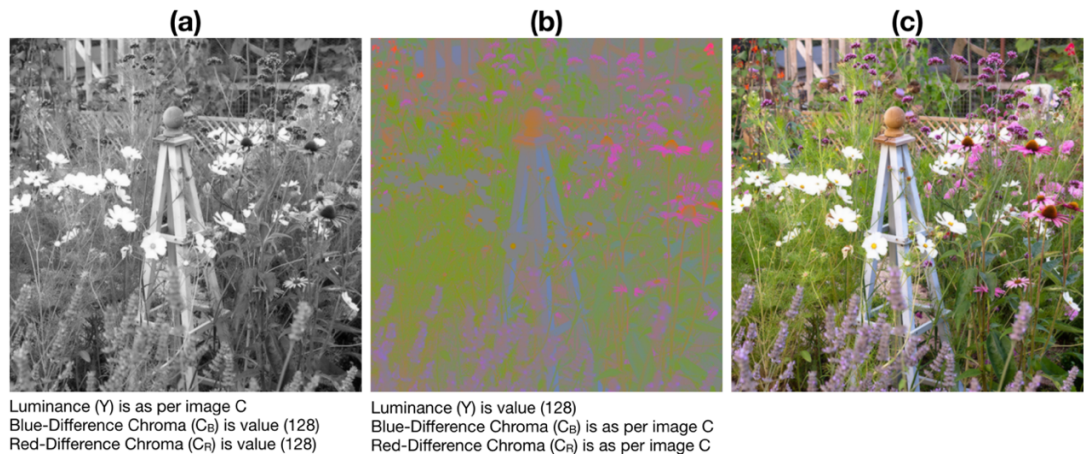


Figure 1.2 Luminance (a) and chrominance (b) components of the Flower Garden image (c) [4] in the 8-bit digital YCbCr colour space.

A variety of psychophysical paradigms yield quantitative data from qualitative judgements of human observers with respect to subjective image quality, or its attributes [5]. Statistical analysis of this data, based on laws of psychophysics, allows it to produce various numerical psychometric scales [5]. It is desirable that such scales are expressed in intervals of just-noticeable differences (JND) for meaningful description of quality levels. However, psychophysical experiments are generally slow, expensive, and difficult to carry out

accurately. They also do not provide easily a causal breakdown of the contributions of different components and processes to the subjective image quality of capture systems.

Image quality metrics (IQM) are often used in conjunction with psychophysical experiments when developing capturing systems. The most suitable metrics originate from imaging and vision science. They map system performance measurements to single figure variables that aim to correlate with psychophysical ratings of image quality (Figure 1.3). Most IQMs implement HVS modelling and account for multiple attributes. Thus, they correlate with subjective image quality more accurately than individual performance measures.

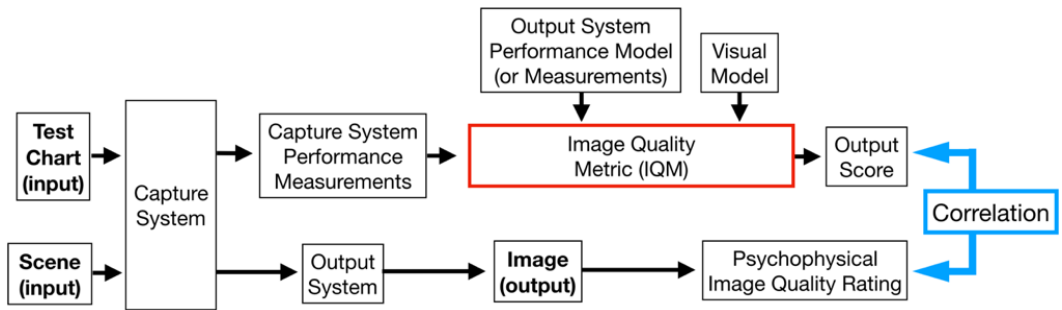


Figure 1.3 Diagram summarising the function of capture system image quality metrics (IQM).

Spatial capture system IQMs generally implement Modulation Transfer Function (MTF) and Noise Power Spectrum (NPS) performance measures which characterise system signal transfer (relating to resolution and sharpness) and noise, respectively. The MTF and NPS are defined in Sections 2.3.2 and 2.3.3, respectively. They are founded upon linear system theory [6, pp. 233–269], and are, strictly speaking, only applicable to linear, spatially invariant and homogenous systems that comply with Equations A1 to A4. Such systems deliver consistent performance regardless of the input signal (Figure 1.4(a)). Spatial capture system IQMs cascade visual contrast sensitivity functions (CSF) with these performance measures as weighting functions. This accounts for whether the measurements are perceptually relevant.

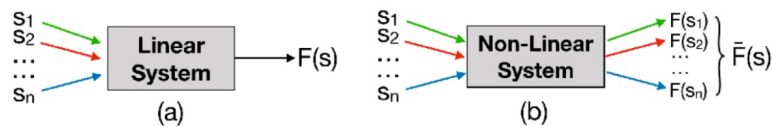


Figure 1.4 Characteristics of system performance measurements, $F(s)$, derived from a range of pictorial scene (or test chart) input signals (s): a) describes a linear system, b) describes a system applying non-linear image signal processing (ISP). $\bar{F}(s)$ is the mean of $F(s_1)$ to $F(s_n)$.

The implementation of advanced ISP algorithms has improved capture system image quality significantly. Denoising and sharpening are two algorithms frequently applied. The simplest, linear denoising filters remove undesirable noise but also detail and edge sharpness that are critical to image quality. Conversely, simple linear sharpening algorithms enhance detail and sharpness but amplify noise. These denoising and sharpening algorithms cancel each other out to some extent when applied together, and image quality is not fully optimised.

It is more common for contemporary capture systems to apply adaptive (or content-aware) denoising and sharpening algorithms. These reduce blurring of the image when denoising, and amplification of noise when sharpening, by filtering image signals selectively; this is described further in Section 2.2.1. These non-linear algorithms optimise image quality more effectively than simple, linear filters and cancel each other to a lesser extent. However, their behaviour is both “*local-content-dependent*” and “*scene-dependent*” (i.e. it varies according to the input signal). Such behaviour is demonstrated in Figures 2.3 and 2.4 and discussed in Section 2.2.1. It follows that the spatial imaging performance of any system applying these algorithms is also scene-dependent, Figure 1.4(b) [2, p. 198], [7]. It should also be noted that the HVS is a highly adaptive system, and visuo-cognitive processes carried out during image quality judgement are expected to be scene-dependent as discussed in Section 2.5.

Spatial capture system IQMs do not account for system scene-dependency [8]. Prior art demonstrates they predict image quality relatively accurately for linear systems, but less accurately for scene-dependent, non-linear systems and/or processes. Their accuracy is limited in the latter case because they are based on performance measurements derived with respect to synthetic signals, not scenes (discussed further in Section 1.1). Likewise, they use CSFs that describe visual sensitivity to unmasked, synthetic signals, not real scenes, as outlined in Section 1.2.

Figure 1.5 provides an example of such inaccuracies when a leading capture system IQM was applied to images of scenes with varying levels of non-linear compression [9, p. 169], [10], [11]. The metric correlated well with the quality of individual scene images. But the correlations for each scene were separated by gains and offsets, resulting in poor overall correlation. These gains and offsets are expected to be due to the scene-dependency of the compression algorithm and human spatial visual sensitivity being unaccounted for.

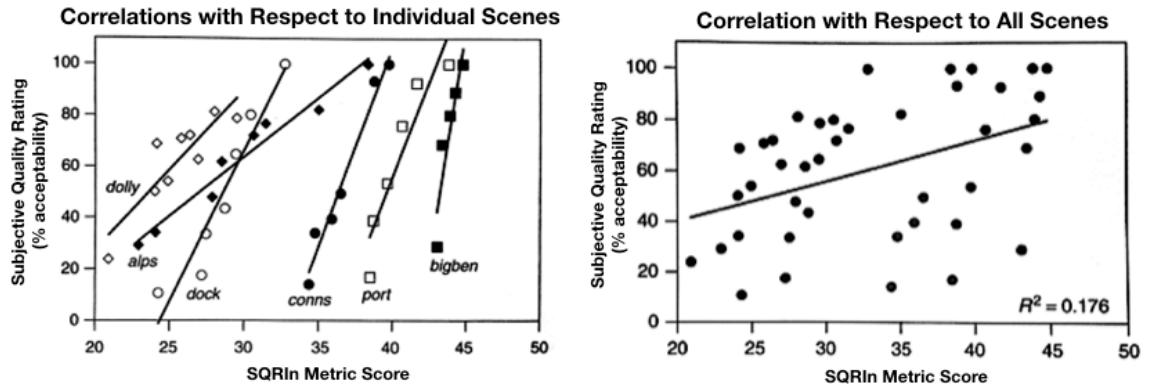


Figure 1.5 Subjective image quality averaged over observer vs the Square Root Integral with Noise (SQRIn) [12] metric at varying levels of non-linear Joint Photographic Experts Group (JPEG) compression. The JPEG quality factor of each image varies from value 100 to where subjective quality drops below 10% acceptability. Adapted from [13]; original data from [10].

This thesis hypothesises, more generally, that the following factors affect the accuracy of spatial capture system IQMs:

- 1) Signal transfer and noise are characterised inaccurately for scene-dependent, non-linear capture systems by their MTF and NPS parameters (discussed in Section 1.1).
- 2) Scene-dependent aspects of lower-level spatial vision, including visual masking, are unaccounted for by their CSF parameters (as described in Section 1.2).
- 3) Cognitive factors of quality judgement are unaccounted for by their CSF parameters.

This thesis revises various capture system IQMs and develops two further, novel IQMs, to investigate whether factors 1) and 2) are significant and can be resolved. Factor 3) has been investigated by the author [14]; it is discussed in Section 2.5. The new IQMs were validated by analysing measurements from simulated camera pipelines. These simulated camera operations in a systematic manner with respect to controlled camera and ISP algorithm variables. Comparable pipelines are used in the industry when developing capture systems.

1.1 Imaging Performance Measures and System Non-Linearity

The MTF and NPS are measured from capture systems using test charts (or test targets). These charts provide a predefined, synthetic, input signal for comparison with the output signal of the capture system. Figure 1.6 shows signals from traditional test charts containing

sine-waves, edges, noise and uniform luminance patches. According to Figure 1.4(a), each test chart should deliver identical measurements for linear systems. However, measurements are derived by different methods for each chart, with different levels of measurement error. Thus, measurements derived using each chart are not identical in practice.

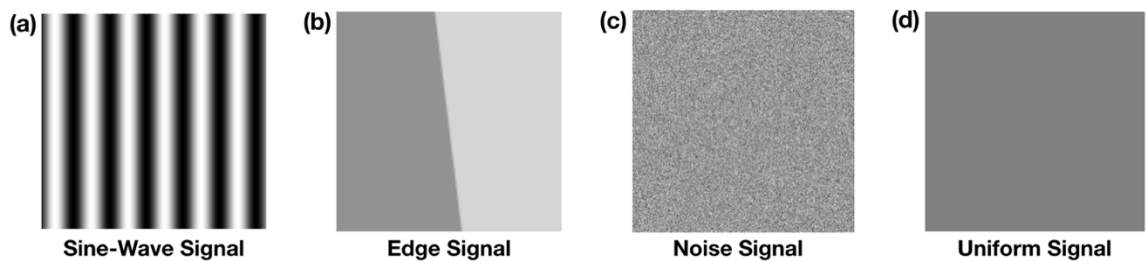


Figure 1.6 Examples of the types of input signals from test charts traditionally used in Modulation Transfer Function (MTF) ((a) to (c)), and Noise Power Spectrum (NPS) measurements (d).

More importantly, for scene-dependent systems that apply non-linear ISP, deriving the MTF and NPS using each of the above test charts delivers different results [2, p. 10], [9, p. 169], [11]. This contradicts Fourier theory [15], and is demonstrated for the MTF by the blue and grey curves in Figure 1.7 that correspond to a slanted edge signal and sinusoidal signal, respectively. Moreover, since each chart is unrepresentative of real scene signals, these measurements describe the average real-world performance of scene-dependent systems inaccurately (i.e. general performance when capturing real scenes).

This problem has been noted by prior art. Performance measures [16]–[21] and IQMs [22] have since been developed for scene-dependent capturing systems that use the dead leaves chart (shown in Figure 2.10(b)); MTFs measured using the dead leaves chart are shown in orange in Figure 1.7. This chart replicates certain statistical properties of the “average” natural scene (as described in Section 2.3.2), making it more representative than the signals in Figure 1.6.

Deriving measurements with respect to images of real scenes, however, is expected to characterise scene-dependent systems more appropriately than measurements from any test chart, provided they are precise and unbiased. This is because such measurements would account for the effect of real scene signals on non-linear ISP algorithms, thus describing performance in real-world image capture scenarios.

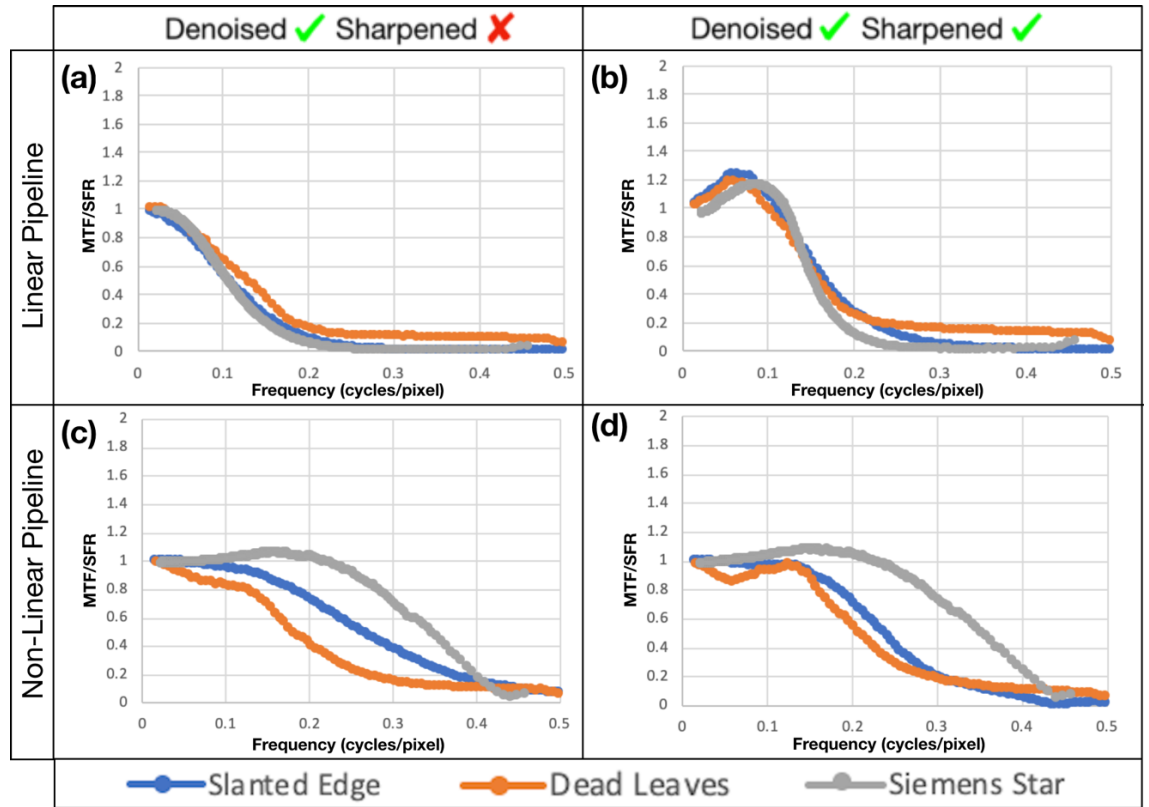


Figure 1.7 MTF measurements obtained from a simulated capture system after linear and non-linear processing, from the following test targets: i) ImatestTM ISO 12233 [23] Slanted Edge Spatial Frequency Response (E-SFR) test chart [24] using sfrmat3 [25] (blue curves). ii) ImatestTM ISO 12233 sinusoidal Siemens Star test chart [26] using Image EngineeringTM analysis software [27] (grey curves). iii) ImatestTM Spilled Coins (dead leaves) test chart using Burns' [28], [29] implementation of the direct dead leaves MTF [19] (orange curves). Both simulation pipelines are described in Section 4.2. The following processes were simulated first: lens blur, Poisson noise at a signal-to-noise ratio of 5, mosaicing and demosaicing [30]. The linear pipeline then applied Gaussian blurring (denoising), and an unsharp mask (sharpening). The non-linear pipeline applied Block Matching and 3D Filtering [31] (denoising) and the Guided Image Filter [32] (sharpening).

Considering Figure 1.4(b), the following measurements are obtained from a non-linear, scene-dependent system if a large number, n , of precise, unbiased performance measurements, $F(s_1)$ to $F(s_n)$, are derived with respect to a set of scene images that are representative of commonly captured scenes, s_1 to s_n :

- 1) $F(s_1)$ in Figure 1.4(b) characterises the performance of the system with respect to a specific input scene image, s_1 , accounting for system scene-dependency.

- 2) $\bar{F}(s)$ in Figure 1.4(b) characterises the average real-world performance of the system with respect to a large number of images of scenes, accounting for system scene-dependency.
- 3) The standard deviation of $F(s_1)$ to $F(s_n)$ in Figure 1.4(b) describes the level of scene-dependent variation in the system's performance (i.e. its scene-dependency).

This thesis develops *scene-and-process-dependent NPS (SPD-NPS) and MTF (SPD-MTF)* measures that characterise system noise and signal transfer, respectively, according to 1) to 3). They are validated by analysing measurements from simulated camera pipelines. Further, they are implemented in a number of IQMs to investigate whether accounting for imaging system scene-dependency improves metric accuracy.

1.2 Human Visual System (HVS) Modelling and Non-Linearity

A wide range of visual models are used in IQMs to describe the function of the HVS. These are reviewed in Section 2.5. Capture system IQMs implement CSFs that describe visual sensitivity to luminance contrast. The CSFs used traditionally in these IQMs have a band-pass shape. They model the reciprocal of the contrast threshold required to detect a sine-wave signal of given spatial frequency on a plain background.

For many years, it has been debated whether these CSFs describe the contrast transduction characteristics of the HVS accurately [33] which are non-linear and locally adaptive. It has also been discussed whether it is appropriate to cascade them with system performance measures founded upon linear system theory [10], [34]. What is clear, is that when observers evaluate the quality of an image, the HVS detects and discriminates complex scene signals and artefacts that are, in general, masked by other content [8]. And that this is a different visuo-perceptive process to the detection of unmasked, narrow-band sine-wave signals.

Recent prior art has developed contextual contrast detection (cCSF) [35] and discrimination functions (cVPF) derived from images of natural scenes to account for HVS non-linearity [8]. Their shape is scene-dependent because they account for the effect of luminance contrast masking on visual sensitivity. They are generally low-pass under most viewing conditions, decaying in high frequencies due to the optical limitations of the eye. These functions were

validated against psychophysical measurements of detection and discrimination of complex scene signals. They are expected to describe the visual processes of image quality evaluation more suitably than the CSFs used traditionally in capture system IQMs [8]. Their implementation in IQMs is investigated in this thesis, with the aim of addressing the hypothesised limitations of current visual models. Previous IQMs have not used them.

1.3 Aims and Overview

The overarching aim of the project is to investigate whether spatial capture system performance can be measured by replacing traditional test charts with images of natural scenes, in order to account for the effects of non-linear, content-aware camera processes. Further, to examine whether such performance measures and contextual human visual models (recently presented in the literature) are more suitable input parameters to spatial image quality metrics.

The objectives of this project are defined below:

- 1) To review and categorise spatial IQMs from a capture system development perspective.
- 2) To examine whether replacing traditional test charts with images of natural scenes produces more accurate measurements of signal transfer (MTF) and noise (NPS) of capture systems that incorporate non-linear content-aware ISP.
- 3) To use revised MTF and NPS measures from 2) and contextual visual models to revise existing IQMs, and to develop novel IQMs that relate to spatial image quality.
- 4) To validate the IQMs from 3) using images from simulated linear and non-linear camera pipelines.

The overview of this thesis is below:

Chapter 2 introduces image quality and fidelity, sources of scene-dependency, spatial image quality attributes, imaging system performance measures, and spatial visual models.

Chapter 3 critically reviews different IQM genres and proposes frameworks for novel and revised IQMs, suitable for camera evaluations.

Chapter 4 proposes and validates three novel NPS measures that characterise noise in capturing systems while accounting for capturing system scene-dependency (SPD-NPS measures).

Chapter 5 proposes and validates three novel MTF measures that characterise signal (contrast) transfer while accounting for capturing system scene-dependency (SPD-MTF measures).

Chapter 6 evaluates a number of variants of the IQMs defined in Chapter 3. These variants use different combinations of the SPD-NPS and SPD-MTF measures, and contextual visual models such as the cCSF and the cVPF.

Chapter 7 presents discussions on the IQM frameworks, SPD-NPS measures, SPD-MTF measures, and the novel and revised IQMs, as well as their implementation and application.

Chapter 8 states the conclusions of the thesis and recommends further work.

1.4 Original Contributions to Knowledge

Research from this thesis demonstrated the following original contributions to knowledge:

- Three novel *Scene-and-Process-Dependent Noise Power Spectrum (SPD-NPS)* measures for characterisation of camera noise, derived either using the dead leaves chart or images of natural scenes.
- Three novel *Scene-and-Process-Dependent Modulation Transfer Function (SPD-MTF)* measures that characterise camera signal transfer, using either the dead leaves chart, or images of natural scenes.
- Successful revision and verification of classical spatial IQMs by substituting existing MTF and NPS parameters with the newly introduced SPD-NPS and SPD-MTF measures, to account for system scene-dependency.
- Development of novel spatial IQMs designed for scene-dependent capture systems (*log Noise Equivalent Quanta (log NEQ)* and *Visual log NEQ*). These elegant and simple IQMs are proven to compete with state-of-the-art but more complicated IQMs.
- Novel measures for quantification of scene-dependency in camera system signal transfer and noise measurements.

Chapter 2 Image Quality and System

Performance

Background theory on image quality and imaging system performance and their multidimensional nature is introduced. Image quality, fidelity and naturalness are defined. Causes for scene-dependent variations in imaging system performance and perceived image quality are discussed. The image quality attributes and their corresponding system performance measures are reviewed, as well as signal-to-noise-based measures. Human visual system (HVS) models used by relevant image quality metrics (IQM) are presented.

2.1 Defining Image Quality and Fidelity

Image quality evaluation should be considered as a scene-dependent, multivariate process [4] with a number of objective and subjective factors [36]. It is difficult to specify uniquely [33], [37]–[39], and no definition has been accepted as universal.

Engel drum [5, p. 1] defines image quality as: *the integrated set of perceptions of the overall degree of excellence of an image*. He describes image quality perception as the multivariate combination of perceptual attributes, referred to as the “nesses”. They include brightness, lightness, sharpness, fineness of detail, texture, graininess or noisiness, contrast, colourfulness, hue, chrominance, naturalness and usefulness [5], [40]–[42]. They form the basis of our language when describing an image’s overall quality and its attributes.

Schade [43] was one of the earliest to measure imaging system quality in the modern way. He defines image quality as: *a subjective judgement made by a mental comparison of an external image with image impressions stored and remembered more or less distinctly by the observer, who allows for a loss of detail in areas too small to be resolved by the eye*. This definition can be extended to describe the output image quality of an imaging system as a function of the input image signal, the imaging system’s performance, the observer’s visual performance, and their preferences regarding the image’s attributes and aesthetics. Figure 2.1 gives a simplified breakdown of these factors, where the image signal flows

approximately from left to right. Imaging system performance can be separated from the other factors, but visual performance and quality consciousness are interlinked [44], [45].

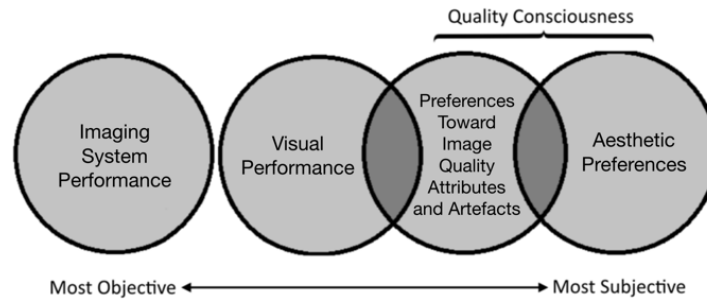


Figure 2.1 Objective and subjective factors that affect perceived image quality.

Keelan [46, p. 9] also considers the practicalities of imaging systems development, defining image quality as: *an impression of merit or excellence, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted*. Inexperienced observers often prefer different weightings of the strength of image quality attributes, compared to more experienced observers [47]–[49]. The former group is preferable, according to Keelan. They are more representative of the general mean of the population (i.e. the potential customer base). Nevertheless, they may overlook subtle artefacts that are noted by experts in rigorous product reviews on websites such as DxOMark™ [50].

Yendrikhovskij [51] defines image quality as: *the degree of apparent suitability of the reproduced image to satisfy the corresponding task*. He elaborates, expressing quality as the multivariate combination of an image’s usefulness, fidelity and naturalness, where their optimal balance differs with the intended application of the image. Usefulness describes an image’s fitness for purpose or the capability of an imaging system to reproduce input scene content in a visibly acceptable way [52]. Fidelity is defined as the *visual equivalence between two images* [34], or more specifically as the perceptibility of differences or distortions between a test and reference image [33], [36], [53]. The reference image is assumed to be ideal, meaning any distortions in the test image reduce fidelity by default [36], regardless of their effect on its perceived quality. However, it is the acceptability of these differences that is relevant, when assessing the relative levels of quality of two images. This requires the quality consciousness (or goodness criteria) of the observer to be accounted for (Figure 2.2).

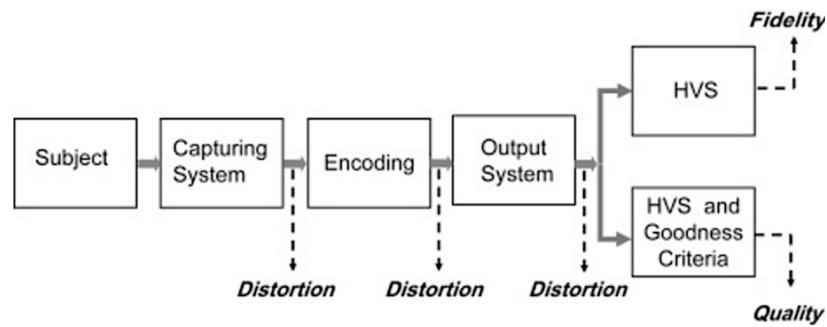


Figure 2.2 Image fidelity and quality evaluation processes with respect to imaging chain distortions [1, p. 351]. Adapted from Ford [9, p. 8], and further from [54, p. 34].

Naturalness is defined as [55]: *the degree of apparent match between the reproduced image and the internal references, e.g. memory prototypes*. These internal references form part of the quality consciousness of the observer and are comparable to the internal references consulted during image quality evaluation [13]. However, they are not identical to the latter. For example, slight over-enhancement of sharpness or contrast can increase quality but decrease naturalness [14].

2.2 Scene-Dependency

Scene-dependency refers to any variation in the performance of an imaging system, the perceived quality of its images, or the perceived strength of its attributes, which is dependent on input scene content. It has been investigated widely [4], [56]–[60] and adds unpredictability to image quality datasets. Scene-dependency with respect to overall perceived image quality increases when test images are of higher quality [57].

It results from scene-dependent variations in:

- 1) objective imaging system performance;
- 2) visual attention patterns and HVS performance;
- 3) observer preference regarding quality attributes and aesthetics [60].

Sections 2.2.1 to 2.2.3 review 1) to 3) above, respectively.

2.2.1 Imaging System Performance

System performance scene-dependency has long been measured, and compensated for, in both analog and digital capturing systems. Certain sources of scene-dependency apply to both types of systems and can be reduced by ensuring that images are captured in focus, with a wide depth of field, at similar camera-to-subject distances, under consistent illumination and without extremes of contrast [60]. While analog (film) capture systems are generally stationary and isotropic [6, p. 210], their tone reproduction is often non-linear (refer to Appendix A). The latter can be compensated for, however, and system performance measures based on linear system theory have been applied successfully to such systems, especially after non-linear tone reproduction [61], [62] and micro-chemical effects [61] were taken into account.

Digital systems, however, employ discrete sensor arrays consisting of finite sampling apertures, causing non-stationary and non-isotropic behaviour [63] (defined in Appendix A). Consequently, the response of the system varies with respect to input signal phase and angle [63], [64]. Modulation Transfer Functions (MTF) measured from different test charts vary with respect to changes in the orientation of the chart, as well as translation [2, p. 197]. The greatest cause of scene-dependency in digital imaging system performance, however, results from implementing non-linear content-aware Image Signal Processing (ISP) such as demosaicing, denoising, sharpening and compression, which are summarised below. These algorithms cause the objective level of image attributes and artefacts to be dependent on local image content [4], [60], [65], thus rendering system performance scene-dependent. This scene-dependency is compounded when multiple algorithms are applied.

Demosaicing interpolates between the pixels of the sparse single-channel images created by Colour Filter Array (CFA) elements, to produce full-colour images according to an RGB additive colour model. Linear demosaicing averages over neighbouring pixels in the spatial [30] or frequency [66] domain. Non-linear content-aware demosaicing is generally a minor cause of system scene-dependency. Non-linear spatial domain methods [67]–[70] are edge-aware. They omit or weight neighbouring pixel content before averaging, depending on local luminance gradients, to attempt to preserve edges and mitigate noise and colour artefacts. Other non-linear methods operate in the wavelet [71] or frequency [72] domain.

The role of denoising is to discriminate between image pixel values that are supposed to be there (i.e. scene content) and spurious image information (noise) [73]. This task is highly challenging, and no denoising algorithm is ideal. Artefacts associated with imperfect denoising include loss of image detail, texture and edge contrast. Linear denoising involves applying mean or Gaussian filters either by convolution or multiplication in the frequency domain. Alternatively, non-linear median filters can be applied. All of these filters ‘average out’ the noise which, incidentally, reduces the intensity of image details and edges.

Non-linear content-aware denoising filters are less detrimental to quality than the above methods but are a significant cause of system scene-dependency. Image structure impedes, in general, the local removal of noise by these filters, as shown in Figure 2.3. Their signal transfer is also dependent on local image structure (Figure 2.4). For many spatial domain content-aware denoising filters, this is caused by applying thresholding before averaging to reduce denoising in the presence of local luminance gradients [74], [75]. This thresholding results in unmasked noise (which is most visible) being filtered heavily while image edges are preserved. But texture and detail are often mistaken as noise and removed.

Further, *Block Matching and 3D Filtering (BM3D)* [31] and Non-Local Means [76] filters operate on a patch-wise level [73]. These content-aware algorithms compare the structure contained within a given small window (e.g. 8-by-8 pixels) with the structure in other windows across the image. Content from the “matching” windows is then averaged with the original window to remove noise. Sumner *et al.* [73] observe that it is impossible to predict the effect of such algorithms on a small-scale image feature in a real-world capture scenario since the actions of the algorithm depend entirely on the surrounding image features which may vary considerably. Other content-aware filters operate in the Discrete Cosine Transform (DCT) [77] or wavelet [78] domains or employ machine learning [79]–[83]; each of the mentioned content-aware denoising filters results in comparable artefacts concerning texture/detail loss.

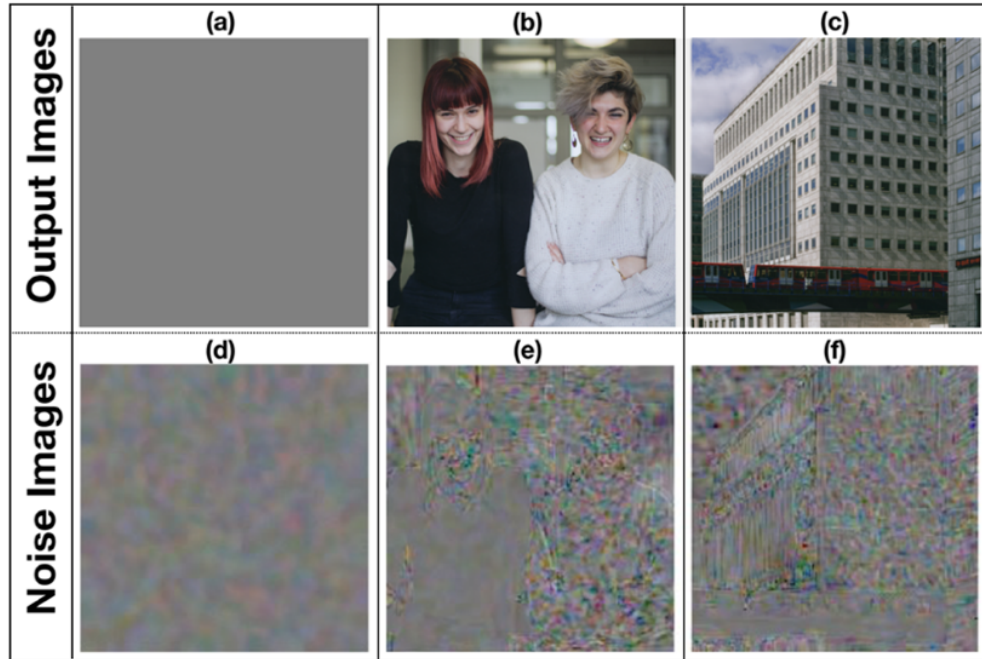


Figure 2.3 Scene-and-process-dependent noise images obtained using ten replicate captures following the method of Section 4.1. All output images and replicates were generated using a simulated image capture pipeline after Poisson noise was added at a linear signal-to-noise ratio (SNR) of 10 and non-linear content-aware denoising [31] and sharpening [84] were applied. Noise images (d), (e) and (f) correspond to the following input information: (a) a uniform-tone patch, (b) the ‘Students’ image [7], and (c) the ‘Architecture’ image [7], respectively (Appendix C). The contrast of each noise image was increased to emphasise scene-dependency.

Sharpening enhances edges, detail and high-frequency contrast but also amplifies noise and other artefacts. Linear sharpening methods include applying high-pass filters or spatial domain unsharp masks (USM) [85]. Non-linear content-aware sharpening causes scene-dependency in imaging system performance because the performed local contrast adjustments are dependent on local image structure (as demonstrated in Figure 2.4). The unintended local amplification of noise (and other artefacts) is also dependent on local image structure since the objective level of noise/artefact amplification relates directly to the level of contrast amplification. For example, non-linear content-aware USMs [86]–[90] lower their amplification of contrast in regions of low local luminance gradient, to attempt to mitigate the perceived amplification of noise, while still enhancing the structural signals most relevant to perceived image quality. Other non-linear content-aware sharpening algorithms operate in various domains [32], [84], [91], [92], providing detail enhancement, edge sharpening and even optional denoising [84], [91]; some employ guidance images [32] or multi-scale contrast manipulations [92].

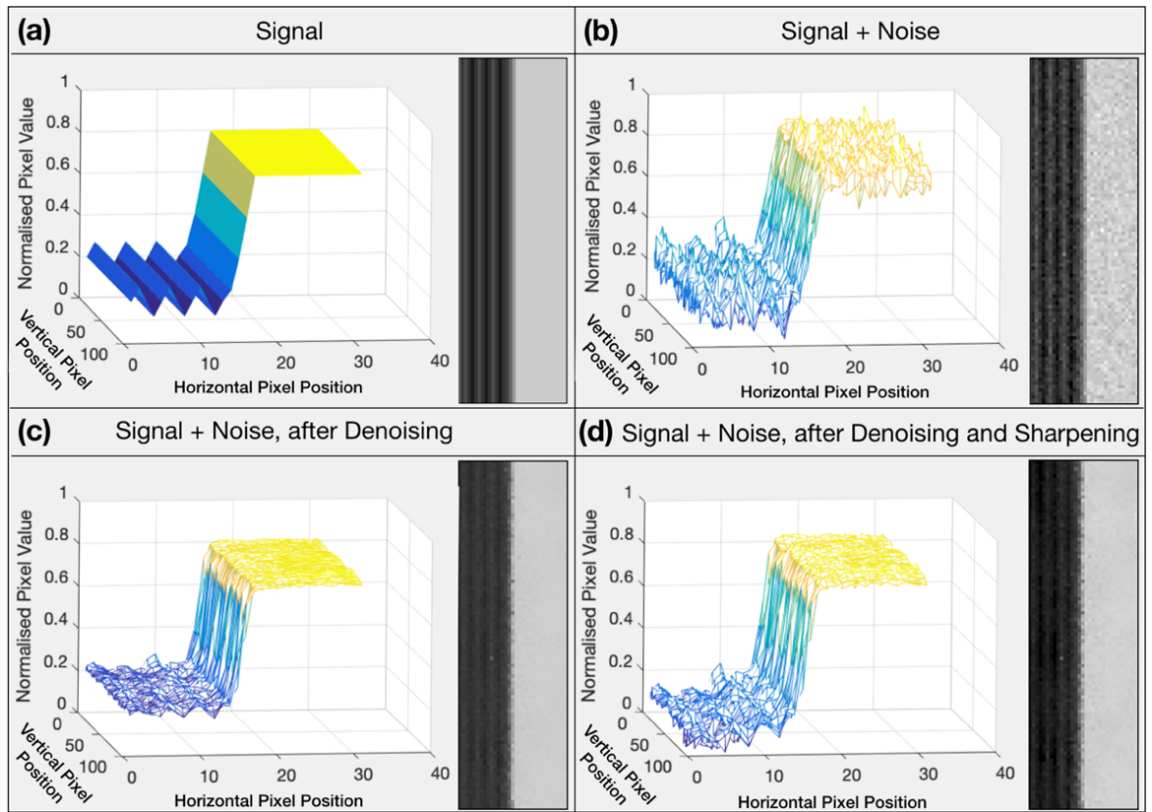


Figure 2.4 Demonstration of local-content-dependency in terms of signal transfer, caused by content-aware denoising and sharpening. Normalised pixel values, y , are plotted vs horizontal, x , and vertical pixel coordinates, z , for: (a) a two-dimensional (2D) input signal consisting of a low-contrast high-frequency signal ($0 < x \leq 15$), high contrast edge ($15 < x \leq 18$), and uniform tone signal ($18 < x \leq 30$). (b) shows (a) after adding Gaussian noise. (c) shows (b) after denoising with the Fast Bilateral Filter (FBF) [93]. (d) shows (c) after sharpening with the Weighted Least Squares (WLS) [92] filter.

Tone mapping converts captured image data to a suitable, output-referred state for viewing on the output device [1, p. 241]. Global, one-dimensional operations (e.g. gamma correction) are suitable if the dynamic range of the subject does not significantly exceed the usable dynamic range of the capture system [94] (which is typically between five and nine exposure stops [1, p. 241]). For subjects with wider dynamic range, implementing local tone mapping algorithms results in improved shadow and highlight detail. These adaptive algorithms map one input pixel value to a number of potential output values, depending on the values of other pixels in the region [94], [95]. Very high dynamic range subjects require high-dynamic-range (HDR) processing, described below, to avoid clipping their shadows and highlights. Multiple, bracketed exposures are captured. A high bit-depth HDR image is constructed (typically 32-bits per channel). The subject's radiance value is estimated at each

pixel location using information from all exposures after it has been weighted according to its reliability [1, p. 242]. The resultant image's bit-depth is reduced to 8 or 16 bits per channel, for output, using image-rendering algorithms that can either be global or local [96] (like the tone-mapping algorithms described above). Content-aware intensity adjustments from local image-rendering and tone-mapping algorithms can affect system signal transfer and noise significantly, in a local-content-dependent and scene-dependent manner.

Compression is applied to reduce the file size of an image by lowering the average bit rate (i.e. the bit allocation per pixel). Image compression can be objectively and perceptually lossless. However, lossy Joint Photographic Experts Group (JPEG) [97], [98] and JPEG 2000 [99] algorithms are applied frequently by image capture systems. Both algorithms are highly non-linear but not content-aware. The perceptual significance of their artefacts is scene-dependent [60], [65]. Significant variation was also found between the MTF of JPEG compression when it was derived from edges and sinusoidal signals [9]–[11], [15]. IQMs that employed MTFs derived from these signals also failed to describe the perceived quality of JPEG compressed images accurately [9]–[11] (Figure 1.5) since they did not account for the algorithm's scene-dependency. This suggests that the scene-dependent characteristics of compression algorithms, and other spatial ISP algorithms such as denoising and sharpening, are relevant to perceived image quality and should be accounted for by spatial IQMs.

2.2.2 Human Visual System

Image quality attributes and artefacts are less noticeable in the presence of certain image content due to the effects of visual masking (defined in Section 2.5) and saliency [60], [65], [100], [101]. Thus, for spatially varying images of scenes, uniformly distributed artefacts are more apparent in certain regions, over others. Likewise, if images of different scenes are generated that contain the same objective level of artefacts, the perceived level of these artefacts often varies between the scenes, as demonstrated by Figure 2.5. The same is also true of their overall perceived image quality. Keelan [46, Ch. 10] describes such variations as scene susceptibility. Saliency relates to the probability of observers fixating their visual attention on a given image location in a free-viewing (or image quality evaluation) scenario.

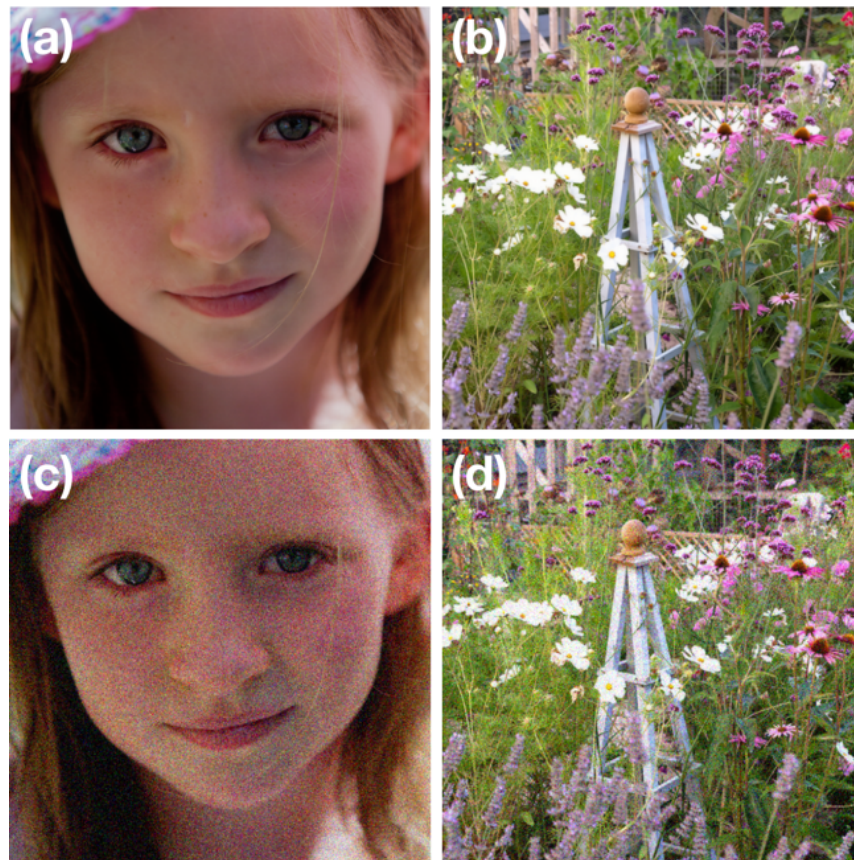


Figure 2.5 Demonstration of human visual scene-dependency: (a) non-busy portrait [102], (b) busy landscape image [102]. (c) and (d) show the same images with Gaussian noise added at identical levels. Visual masking suppresses noisiness and perceived quality loss due to noise in (d).

The factors that affect visual masking are the viewing conditions, the frequency content of the attributes/artefacts in question, and the frequency content of local masking signals or noise. Attributes/artefacts are also more noticeable if they are located in salient image regions which attract visual attention. This includes positions near to the foreground, prominent subjects, text, distinct edges, luminance/colour contrasts, or objects that deliver the meaning of the image. Visual attention patterns can be affected by the viewing task and the level of experience of the observer [45] and can be predicted by certain algorithms, as shown in Figure 2.6.



Figure 2.6 Visual saliency maps predicted by the Graph-Based Visual Saliency algorithm [103]. Adapted from [41, p. 43].

2.2.3 Observer Preference

The aesthetic preferences of observers and their tolerance toward image quality attributes and artefacts are dependent on scene content [104], [105] and can vary with experience [47]–[49] and the viewing task [45]. They are also affected by mood and state of awareness [106], as well as ethnicity and cultural factors [107]. For example, observers prefer skin textures to be “softly focused” compared to other image textures and signal content such as edges. Consequently, close-up portraits are often preferred to be less objectively sharp than other types of images, such as landscapes [104], [105]. Further, the subjective quality of the former decreases when objective quality (sharpness) increases past a threshold level [105]. The aesthetic preferences of observers are affected by the distribution of objects in images [108] and their interrelatedness [109]. For example, related objects are preferred to be positioned closer together than unrelated objects [109].

2.3 Image Quality Attributes

The visual attributes of spatial image quality are defined (contrast/tone, resolution, sharpness and noise). Resolution, contrast and noise are also discussed as physical attributes of imaging systems and images. The objective system performance measures that are used to describe each of these attributes are also defined in this section.

The final attribute, colour, is defined objectively (and subjectively) by the reproduction (and perception) of image brightness/lightness, chroma/colourfulness/saturation and hue [1, p. 346]. Brightness and lightness also relate to spatial attributes of contrast/tone and are defined in Section 2.3.1. The other colour attributes are not defined, however, because this project is concerned with spatial IQMs and system performance measures only. The IQMs and performance measures developed in this thesis are validated using colour images. However, the ISP algorithms used by the simulation pipelines affect spatial image quality attributes, in particular, producing mainly spatial artefacts.

The quality attributes form logical components of the observer image quality judgement process [57]. They are fundamental to image quality analysis [13], [38], [52], and the language used to describe relevant perceptions. Their presence can be preferential or detrimental (artefactual) to the overall perceived image quality [46, p. 7], [56, p. 1], [110, p. 1] or preferential up to a point and detrimental thereafter [14]. Attributes that are perceived to be of the highest magnitude have a disproportionate influence on the overall perceived image quality [110], [111]; they suppress attributes of lower perceived magnitude [56, p. 4] and become identifiable features of an image (e.g. a “sharp” image). This is particularly the case for low-quality images [112] for which isoquality contours are particularly steep edged [111].

Of interest to this thesis are the MTF (Section 2.3.2), Noise Power Spectrum (NPS, Section 2.3.3) and Noise Equivalent Quanta (NEQ, Section 2.4) measures of imaging system signal transfer (relating to resolution and sharpness), noise, and signal-to-noise performance, respectively. These measures are based upon linear system theory (defined in Appendix A) and are transformed into the spatial frequency domain. This allows measurements for each imaging chain component to be cascaded at different spatial frequencies, thus expressing the performance of the entire imaging chain. Moreover, if these measurements are projected at the plane of the observer’s eye, they can be weighted with the human visual contrast

sensitivity function (CSF) to emphasise the most visually important information [41, p. 6]. This makes them particularly useful input parameters for IQMs, as discussed in Chapter 3.

2.3.1 Contrast and Tone

Subjective contrast is defined as perceived luminance variation [35]. Likewise, colour contrast refers to perceived chrominance variation. Objective differences in luminance (and chrominance) signals across the surface of a two-dimensional (2D) still image produce visual sensations of contrast (and colour contrast), as shown in Figure 2.7. Spatially varying luminance contrast is the principal carrier of image information [33], with respect to perceptions of image structure [113]. Scene structure is fundamental to visual understanding [114]–[116], making luminance contrast core to perceived image quality and fidelity, as well as four out of the five quality attributes: resolution, tone, sharpness and noise. This is exemplified by Figure 1.2(a) which is visibly sharp, contains tonal information, and is far more recognisable and detailed than Figure 1.2(b). Chromatic contrast signals are accounted for by many specialist capture system colour metrics and IQMs. However, they are overlooked by a large number of spatial IQMs, including those developed in this thesis. Consequently, this section focuses mainly on luminance contrast.

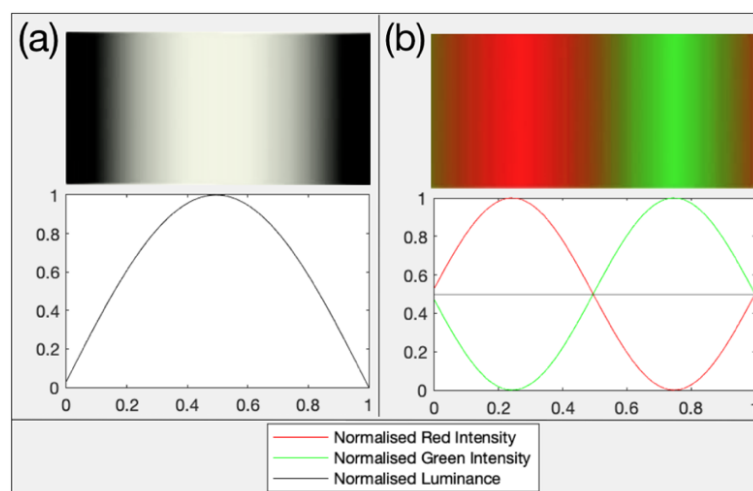


Figure 2.7 Sinusoidal modulations in terms of: a) luminance, and b) red-green chrominance signals; adapted from [117].

Subjective perception of luminance contrast in broadband images is a non-linear process [118] that compares local variations in luminance with the surrounding luminance. Adaptations of human vision to luminance and chrominance are also affected by local signal

intensities and the viewing conditions [119], [120]. They can lead to simultaneous contrast phenomena [1, p. 96], [38] that biases both contrast and colour perception [1, p. 96].

Weber contrast, C_{Weber} , is the simplest contrast measure (Equation 2.1). ΔL is the difference between the luminance of a single small stimulus and the uniform background luminance, L [1, p. 338], [118]. Contrast perception is shown to relate to relative changes in luminance, not absolute changes [121, Ch. 8.3]. The perceptual relevance of Weber contrast decreases, however, when stimuli increase in size or number across the visual field [118], making it a poor measure of pictorial image contrast.

$$C_{Weber} = \frac{\Delta L}{L} \quad (2.1)$$

Michelson contrast, $C_{Michelson}$, describes suitably the contrast (or modulation) of predictable periodic functions such as sine-waves where the distribution of higher and lower luminances are of equal proportional area (Equation 2.2) [6, p. 212], [118]. L_{Max} and L_{Min} are the maximum and minimum stimulus luminances, respectively. It is an unsuitable measure for pictorial scene contrast because all other signal intensities are unaccounted for [113], [122].

$$C_{Michelson} = \frac{L_{Max} - L_{Min}}{L_{Max} + L_{Min}} \quad (2.2)$$

Root mean square (RMS) contrast, C_{RMS} , accounts for all signal intensities (Equation 2.3). It is thus a more appropriate measure of pictorial image contrast and is employed widely in visual and image quality research for its simplicity. For an image of dimensions $M \times N$, L_{ij} is the luminance of the i^{th} and j^{th} pixel normalised to $0 < L < 1$. \bar{L} is the mean luminance. The RMS contrast spectrum, or simply the contrast spectrum, refers to the RMS contrast of isolated image spatial frequency bands versus their spatial frequency. It is used in the computation of various visual models defined in Section 2.5.

$$C_{RMS} = \sqrt{\frac{1}{MN - 1} \sum_{i=1}^M \sum_{j=1}^N (L_{ij} - \bar{L})^2} \quad (2.3)$$

RMS contrast has two limitations as a perceptual contrast measure for pictorial images. It applies equal weighting to all spatial distributions of image contrasts [113]. Also, the spatial frequency of image contrasts is unaccounted for [118], which affects their perceived contrast

magnitude [123], [124]. Local Band-Limited (LBL) contrast addresses these limitations by accounting for HVS quasi-local processing and is computed as follows [118], [123]. Images are filtered to yield several single-octave spatial frequency bands. The LBL contrast, $C_{LBL}(x, y)$, for each band, $a(x, y)$, at pixel coordinate (x, y) , is given by Equation 2.4 [118], [123]. $l(x, y)$, is a local luminance mean image containing the energy of the image at all frequencies below the band in question. Despite the intricacies of contrast perception in pictorial scenes, Triantaphillidou *et al.* [125] found the mean of the RMS contrast correlated well with the mean of the integral of the LBL contrast, across a large number of natural scene images, at relevant spatial frequencies.

$$C_{LBL}(x, y) = \frac{a(x, y)}{l(x, y)}, \quad \text{where } l(x, y) > 0 \quad (2.4)$$

Tone reproduction is the most critical component of image quality and fidelity and influences perceived sharpness, visual resolution and colour significantly [1, p. 377]. It is defined objectively as the relationship between the intensities and intensity differences of a given scene and any reproductions of that scene [38]. Subjectively, tone reproduction refers to the observer's impression of these characteristics and relates to perceptions of scene contrast, lightness or brightness [1, p. 377]. Brightness is defined as the degree to which a signal appears to display more (or less) light [1, p. 78]. Lightness is a relative form of brightness, where the brightness of a given signal is judged relative to the brightness of a white signal [1, p. 78]. Increasing global luminance contrast [126], [127] or the contrast ratio [128] can increase perceived sharpness. It can also raise perceived quality provided that the signal-to-noise ratio (SNR) does not decrease and clipping or display non-linearity does not occur [126].

Imaging system tone reproduction is characterised by tone transfer functions concerning input-to-output intensity. ISO 14524 [129] defines the Opto-Electronic Conversion Function (OECF) that is the tone transfer function for digital image capture systems. Input luminances are plotted against output pixel values, as measured from a TE264 or equivalent chart [130] (Figure 2.12). Plotting measurements of the reverse process characterises the Electro-Optical Conversion Function (EOCF) of a display. The tone reproduction of a capture-display imaging chain is obtained by cascading the OECF, EOCF, and the tone transfer functions of relevant ISP algorithms.

Gamma is a simple measure of global contrast reproduction. It is calculated as the gradient of the relative intensity transfer function of the system (or component) when this function is expressed in a log-log space, or as the exponent of the same transfer function expressed in a linear-linear space [1, p. 378]. Tonal distortion can be prevented by correcting for the gamma of individual components or the imaging chain. The gamma of an imaging chain is given by the product of the gamma of its components. It is set commonly to between 1 and 1.5 for normal and dark viewing conditions, respectively, to achieve optimal subjective tone reproduction [1, p. 379] where lightness is perceived as linear [9].

Gain Offset Gamma (GOG) models [131] describe system tone transfer functions. For example, the display EOCF of the popular standardised RGB (sRGB) [132] colour space can be approximated by Equation 2.5 where the normalised output luminance and digital input count are V_{sRGB} and V'_{sRGB} , respectively, gamma, γ , is 2.2, gain, α , is 1, offset, β , is 0. The white point of the reference display is D65 with CIE xy chromaticity values of 0.3127 and 0.3290, respectively [132, p. 9] and luminance level of 80 cd/m².

$$V_{sRGB} = \alpha(V'_{sRGB} + \beta)^\gamma \quad (2.5)$$

2.3.2 Sharpness and Resolution

Sharpness is a subjective attribute that relates to the visual definition of image edges and texture [1, p. 347]. Objective sharpness is defined as the capability of a system to reproduce object boundaries and edges acutely [16]. This depends on its ability to reproduce contrast, especially at higher spatial frequencies [1, p. 443]. The sharpness of a given image is affected more by the amplitude of its edges/structure than the number of edges or density of the structure [133]. Consequently, low depth of field images can still be perceived as sharp. Blur is the inverse of sharpness.

Image sharpness influences overall impressions of image quality significantly [2, p. 2], [16], [56, Sec. 4.3], [111], [134], [135]. Sharpness and quality correlate well without the presence of image noise [136], but this relationship can break down under certain conditions [52]. The perception of image quality with respect to changes in sharpness is also generally more scene-dependent [60], [137] than the perception of sharpness itself [138], [139]. Image sharpening generally improves perceived image quality, but over-sharpening reduces quality and naturalness [14], [138] in a scene-dependent fashion [60]. Moderate increases in global

contrast and even noise can increase sharpness under some conditions [140]. Evidence of sharpening and over-sharpening are indicated by ISO 12233 [23] slanted edge MTFs reaching above 1 and 1.4, respectively [141].

Resolution is an objective attribute relating to the capability of a system to reproduce fine detail [142]. Visual resolution is a subjective attribute concerning the perceived rendering of detail. Resolution is the most significant factor affecting sharpness [140]. The most critical spatial frequencies regarding sharpness perception occur approximately two octaves above the peak of the standard CSF [14], [139] (refer to Section 2.5 for further information). Thus, reproduction of higher perceivable frequencies (or microcontrast) are important [1, p. 443], [138], [143]. Other research stresses the influence of signals at all visible frequencies [14], [138], [143] or lower frequency structural signals [144] to sharpness.

Early resolution measures are given in line pairs/mm and measured from test charts consisting of equally spaced sharp-edged bars of increasing frequency, such as the USAF 1951 Resolution Test Chart [145]. The spatial limit of resolution is defined by the distance between the centre of two edges when the peak of the Airy disk from one edge coincides with the first minimum of the Airy disk of the other [146]. The reciprocal of this distance is the resolving power [1, p. 35]. Both measures correlate with sharpness to a limited extent, as demonstrated by Figure 2.8 [147]; they are not employed in current IQMs.



Figure 2.8 Heynacher & Kober's [147] resolving power versus sharpness example: (a) has higher resolving power, (b) has higher contrast at mid-high frequencies and is perceived to be sharper than (a) at most viewing distances.

More comprehensive and perceptually relevant measures characterise the transfer of signal modulation (contrast) by systems, with respect to all relevant spatial frequencies. When a system satisfies the requirements of linear system theory, given by Equations A1 to A4, Dainty and Shaw [6, pp. 192–207] state that its signal transfer is defined fully by its Point Spread Function (PSF). Thus, the PSF and the related Line Spread Function (LSF) and Edge Spread Function (ESF) are fundamental to spatial image quality modelling [52]. They are derived by Equations 2.6 to 2.18.

To derive the PSF, the relationship between the input, $f(x, y)$, and output, $g(x, y)$, of an imaging system, $S\{ \}$, is specified first [6, p. 205]:

$$g(x, y) = S\{f(x, y)\} \quad (2.6)$$

The Dirac delta function, $\delta(u)$ [6, p. 192], describes a theoretical point intensity source of infinitesimal width and unit area. Equation 2.7 [6, p. 192] describes its sifting property.

$$\int_{-\infty}^{\infty} \delta(u) f(a - u) du = \int_{-\infty}^{\infty} \delta(a - u) f(u) du = f(a) \quad (2.7)$$

$$\text{where } \delta(u) = 0 \quad \text{for } u \neq 0; \quad \text{and } \int_{-\infty}^{\infty} \delta(u) du = 1$$

Considering Equation 2.7, the input, $f(x, y)$, and output, $g(x, y)$, of Equation 2.6 can be expressed as a linear combination of weighted and displaced Dirac functions [6, p. 206].

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, y_1) \delta(x - x_1) \delta(y - y_1) dx_1 dy_1 \quad (2.8)$$

$$g(x, y) = S \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, y_1) \delta(x - x_1) \delta(y - y_1) dx_1 dy_1 \right\} \quad (2.9)$$

Provided that the system is linear and homogeneous (i.e. Equation A3 is satisfied) the output, $g(x, y)$, can be re-expressed with the operator, $S\{ \}$, inside the integral [6, p. 206]. Thus, the system is described as acting on the Dirac functions, $S\{\delta(x - x_1) \delta(y - y_1)\}$, and the input, $f(x_1, y_1)$, is applied as a weighting function.

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, y_1) S\{\delta(x - x_1) \delta(y - y_1)\} dx_1 dy_1 \quad (2.10)$$

The response of the system at coordinates (x, y) to the Dirac function at coordinates (x_1, y_1) is the PSF, $h(x, y; x_1, y_1)$ [6, p. 207]:

$$h(x, y; x_1, y_1) = S\{\delta(x - x_1) \delta(y - y_1)\} \quad (2.11)$$

Substituting the PSF, $h(x, y; x_1, y_1)$, in place of $S\{\delta(x - x_1)\delta(y - y_1)\}$ in Equation 2.10, expresses the system fully by the PSF at a given spatial location [6, p. 207]:

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, y_1) h(x, y; x_1, y_1) dx_1 dy_1 \quad (2.12)$$

Provided that the system is stationary (i.e. Equation A4 is satisfied) the output can be expressed as [6, p. 207]:

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, y_1) h(x - x_1, y - y_1) dx_1 dy_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - x_1, y - y_1) h(x_1, y_1) dx_1 dy_1 \end{aligned} \quad (2.13)$$

Equation 2.13 is the imaging equation. It states that the output image can be reproduced by summation of the scaled PSFs across the surface of the image. This is equivalent to convolving the input signal, $f(x, y)$, with the PSF, $h(x, y)$, [1, p. 127], [6, p. 207]:

$$g(x, y) = f(x, y) \otimes h(x, y) \quad (2.14)$$

The LSF is defined by Dainty and Shaw [6, p. 209] as the response of the system to an ideal, infinitesimally thin line input, represented by a single delta function, $\delta(x_1)$, on the y_1 axis. Considering the convolutional relationship in Equation 2.14, the LSF, $l(x)$, is given below:

$$l(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x - x_1) h(x_1, y_1) dx_1 dy_1 \quad (2.15)$$

The sifting property of the delta function allows this to be rewritten below [6, p. 209]. Thus, integrating the PSF with respect to one variable yields the LSF.

$$l(x) = \int_{-\infty}^{\infty} h(x, y_1) dy_1 \quad (2.16)$$

The ESF, $e(x)$, is defined using the imaging equation, as the response of the system to an ideal edge signal, $f(x)$, of infinitely high gradient [6, p. 210]. Note that integrating the LSF with respect to x_1 yields the ESF.

$$e(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - x_1) h(x_1, y_1) dx_1 dy_1 = \int_{-x}^{\infty} l(x_1) dx_1, \quad (2.17)$$

$$\text{where } f(x) = 0 \text{ for } x < 0, \text{ and } f(x) = 1 \text{ for } x \geq 0$$

Thus, differentiating the ESF yields the LSF [6, p. 211]. This relationship is useful since practical measurement of the ESF is more straightforward and less error-prone than the PSF or LSF. This is because producing a physical target with an infinitely small point or thin line is impossible.

$$l(x) = \frac{d}{dx}(e(x)) \quad (2.18)$$

The MTF is the standard objective measure for imaging system resolution and sharpness. It characterises the transfer of modulation (contrast) signals by an imaging system, or its components, with respect to spatial frequency. The MTF provides useful information for the prediction of perceived imaging system resolution, contrast, and sharpness. It is implemented in most spatial capture system IQMs. The popular MTF50 objective sharpness metric and the MTF10 metric for the limit of resolution are given by the lowest frequencies where 50% and 10% modulation transfer occur, respectively; MTF10 relates to the Rayleigh Criterion [146]. Both these objective system performance metrics have limitations since they do not account for the reproduction of all relevant frequencies; neither accounts for the HVS.

Dainty and Shaw [6, pp. 211–213] define the MTF and the related Optical Transfer Function (OTF) with respect to a theoretical sinusoidal input signal, $f(x)$, where u is spatial frequency and ε is phase. The amplitude, b , and mean (DC) level, a , of the sinusoid are given by the numerator and denominator of Equation 2.2, respectively.

$$f(x) = a + b \cos(2\pi ux + \varepsilon) \quad (2.19)$$

The output of the imaging system, $g(x)$, with respect to the sinusoid is given by substituting the full description of $f(x)$ into the imaging equation, given by Equation 2.13 [6, p. 212].

$$g(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a + b \cos(2\pi u(x - x_1) + \varepsilon)) h(x_1, y_1) dx_1 dy_1 \quad (2.20)$$

Integrating with respect to the y_1 dimension expresses the output, $g(x)$, in terms of the LSF of the system, $l(x_1)$ [6, p. 212].

$$g(x) = \int_{-\infty}^{\infty} l(x_1) (a + b \cos(2\pi u(x - x_1) + \varepsilon)) dx_1 \quad (2.21)$$

If the LSF is normalised to unit area and Equation 2.21 is expanded according to the trigonometric identity $\cos(A - B)$, $g(x)$ may be re-expressed as [6, p. 212]:

$$g(x) = a + b \cos(2\pi ux + \varepsilon) \int_{-\infty}^{\infty} l(x_1) \cos(2\pi ux_1) dx_1 + b \sin(2\pi ux + \varepsilon) \int_{-\infty}^{\infty} l(x_1) \sin(2\pi ux_1) dx_1 \quad (2.22)$$

This can be expressed in terms of the real, $C(u)$, and imaginary parts, $-S(u)$, of the OTF (Equation 2.23) [6, p. 212]. The OTF, denoted by $OTF(u)$, is thus given by the Fourier transform of the LSF, $l(x_1)$, and is defined by Equation 2.24 [6, p. 212].

$$g(x) = a + b \cos(2\pi ux + \varepsilon) C(u) + b \sin(2\pi ux + \varepsilon) S(u), \quad (2.23)$$

where

$$OTF(u) = C(u) - i S(u) = \int_{-\infty}^{\infty} l(x_1) e^{-2\pi i u x_1} dx_1 \quad (2.24)$$

If the modulus and phase of the $OTF(u)$ are defined as $MTF(u)$ and $\phi(u)$, respectively, then Equation 2.23 can be reduced, as follows [6, p. 212]:

$$g(x) = a + MTF(u) b \cos(2\pi ux + \varepsilon + \phi(u)) \quad (2.25)$$

Comparing Equations 2.25 and 2.19 shows that for a linear and stationary system (as defined in Appendix A) the output, $g(x)$, is a sinusoid of the same frequency as the input. The output modulation is defined by $MTF(u) \frac{b}{a}$ [6, p. 213].

Thus, the MTF can either be defined by:

- 1) Equation 2.26 as the ratio of the output, $M(u)_{output}$, to the input modulation, $M(u)_{input}$, of sinusoidal input signals, where modulation is calculated using Equation 2.2 [6, p. 213];

- 2) the modulus of the Fourier transform of the LSF, $l(x_1)$, as defined by Equation 2.27 [1, p. 133], [6, p. 213].

$$MTF(u) = \frac{M(u)_{output}}{M(u)_{input}} \quad (2.26)$$

$$MTF(u) = \left| \int_{-\infty}^{\infty} l(x_1) e^{-2\pi i u x_1} dx_1 \right| \quad (2.27)$$

Cascading the MTFs of n components according to Equation 2.28 yields the MTF of a full imaging chain, $MTF(u)_{chain}$, [1, p. 133].

$$MTF(u)_{chain} = MTF(u)_i \times MTF(u)_{i+1} \times MTF(u)_{i+2} \dots \times MTF(u)_n \quad (2.28)$$

The PSF, LSF, ESF, OTF and MTF relate to one another mathematically as described below.

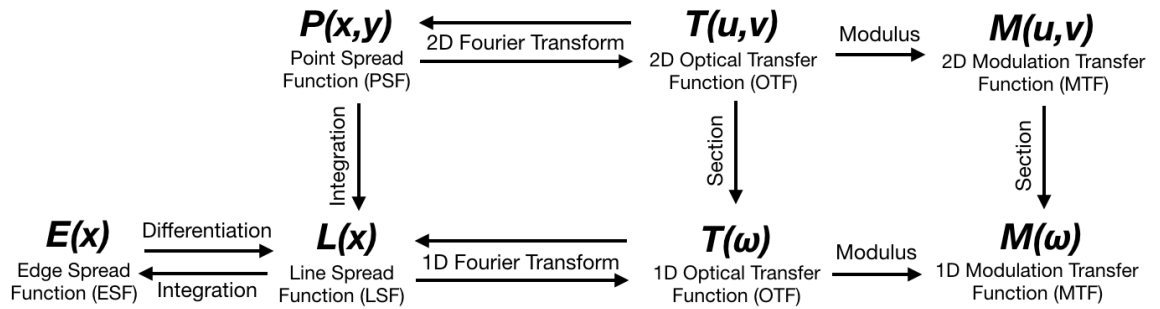


Figure 2.9 Relationships between the various spread functions and transfer functions that describe system signal transfer. u and v are spatial frequencies with respect to the x and y dimensions of the Point Spread Function (PSF), respectively. ω is spatial frequency with respect to the x dimension of the Line Spread Function (LSF) or Edge Spread Function (ESF). Adapted from [1, p. 133].

There are three traditional methods of deriving the MTF. Each employs a different input signal. Firstly, the ESF is measured from an appropriate “ideal” edge signal and differentiated to yield the LSF (Equation 2.18) [148]–[150]. The MTF is then derived using Equation 2.27. The ESF is traditionally measured from a trace of pixels perpendicular to an imaged edge that is aligned with the pixel grid [2, p. 20]. The ISO 12233 Spatial Frequency Response (SFR) method [23], [141] is more popular, however. It uses a slanted edge (typically 5° , as shown in Figure 2.10(a)) which enables supersampling, greatly reducing aliasing. The frequency content of the test chart is not corrected for, but it is assumed to be

constant and equal to 1 at the spatial frequencies of interest. Thus, the measure is often referred to as SFR, not MTF. It describes sharpness well in non-linear systems [146] and is implemented in the Edge Sharpness Metric of the IEEE P1858 Camera Phone Image Quality (CPIQ) standard [22, p. 18]. It is sensitive to noise since relevant signals are locally concentrated around each edge [17]. Edge signals are generally sharpened more and denoised less than natural scene signals by non-linear ISP, as discussed in Section 2.2.1.

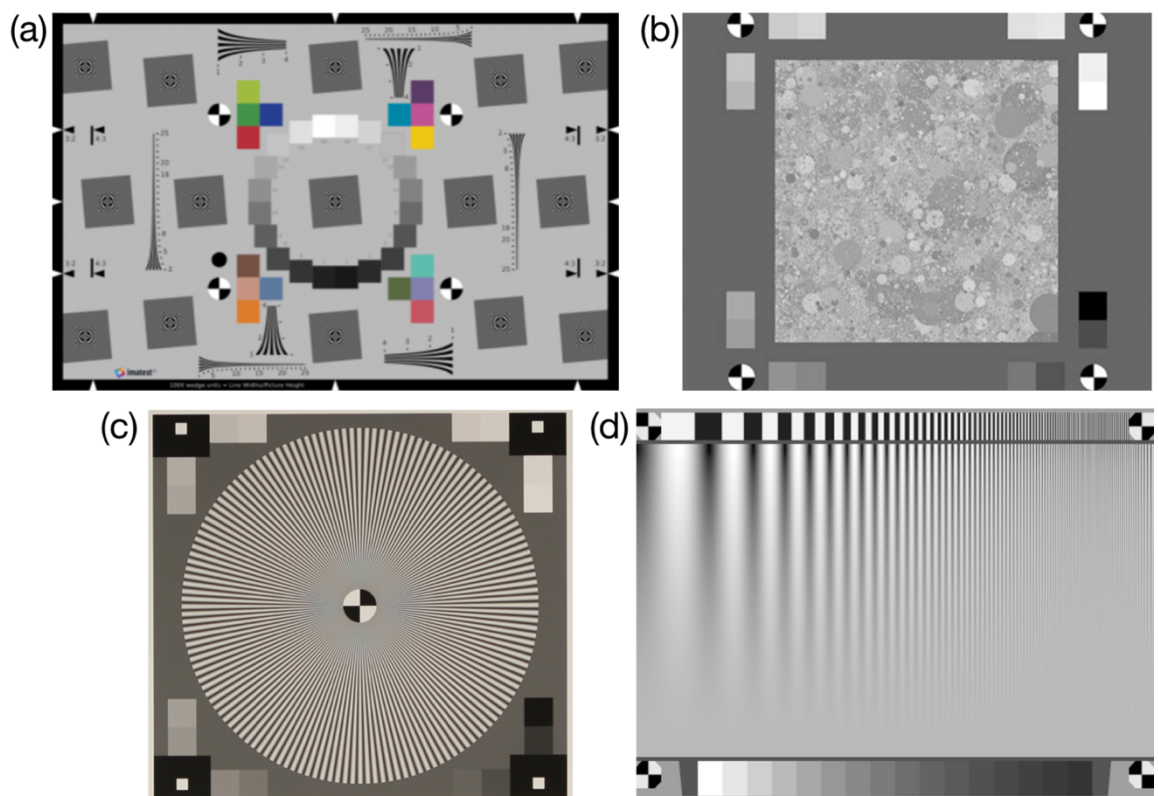


Figure 2.10 Test charts used to derive measurements of capture system signal transfer: (a) Imatest™ ISO 12233 [23] Slanted Edge Spatial Frequency Response (E-SFR) test chart [24]. (b) Image Engineering™ TE265 Dead Leaves test chart [151]. (c) Imatest™ Siemens Star test chart [26]. (d) Imatest™ Log Frequency-Contrast test chart [152].

Alternatively, the MTF is measured from sinusoidal input signals using Equation 2.26, by comparing output to input modulation, $M(u)$, where both are given by Equation 2.2. Measurements can be derived from several sine-wave patches of single frequency [150], [153]–[155] (e.g. Figure 1.6(a)), the ISO 12233 Siemens Star Chart [141], [142] (Figure 2.10(c)) or single sinusoids of varying frequency [2, p. 27], [141], [152] (Figure 2.10(d)). MTFs measured from sinusoidal signals describe system limiting resolution well since they respond less to sharpening [146].

Finally, following Equation 2.38, the MTF can be derived by comparing output and input one-dimensional (1D) luminance power spectra, $PS(u)$, of a white noise signal [150] in Equation 2.29, where $PS(u)$ is given by the rotational average of $PS(u, v)$ in Equation 2.30.

$$PS(u)_{output} = MTF(u)^2 \cdot PS(u)_{input} \quad (2.29)$$

The 2D luminance power spectrum or power spectral density (PSD), $PS(u, v)$, for discrete systems is given below [29]. $I(x, y)$ is a luminance image of dimensions $M \times N$. u and v are spatial frequencies with respect to x and y . In this thesis the term power spectrum refers to the spectrum of the signal. The term NPS refers specifically to the power spectrum of noise, defined in Equation 2.36.

$$PS(u, v) = \left| \sum_{x=\frac{M}{2}+1}^{\frac{M}{2}} \sum_{y=\frac{N}{2}+1}^{\frac{N}{2}} I(x, y) e^{-2\pi i(ux+vy)} \right|^2 \quad (2.30)$$

For linear systems, the MTF is a unique property of the system, in theory, and deriving it using any of the above measurement methods from edges, sinusoidal signals or noise should fully specify the system. Thus, for an idealised, noiseless, linear and stationary system, measurements delivered by each of these methods should converge, as described by $F(s)$ in Figure 1.4(a).

However, in practice, digital imaging systems are non-stationary, and the MTF varies depending on whether the target signal is in or out of phase with the sampling array [156]. MTFs measured by each of the above methods also produce different levels of measurement error (i.e. variation error and bias) [2, p. 198], [41, p. 22], [157], [158].

These errors result from:

- 1) inaccurate specification of the input (test chart) signal;
- 2) inaccurate measurement of the output signal from the system;
- 3) influence of imaging system noise;
- 4) “measurement noise” resulting from processing of input (1) and output (2) signals when computing the MTF (e.g. discrete Fourier transform (DFT) computation).

Consequently, although linear digital systems can be said to have a unique MTF, it cannot be obtained in practice with a single measurement. Averaging over several measurements can result in some convergence toward it by mitigating the effect of measurement noise [157]. Nevertheless, strictly speaking, each of the above methods delivers a slightly different MTF in practice for such systems, none of which are capable of fully specifying the system.

More importantly, when measurements are derived from systems implementing non-linear content-aware denoising, sharpening and tone-mapping ISP, as well as lossy compression, the differences between MTFs obtained from each of the above methods are compounded significantly [146], [159] (as demonstrated in Figure 1.7, and illustrated by $F(s)$ in Figure 1.4(b)). This is because these ISP algorithms react differently to input edges, sinusoids and random noise signals (see Section 2.2.1 for further information). This causes local-content-dependent variations in system signal transfer (and noise) that render the MTF (and NPS) of the system to be *target-dependent* (i.e. dependent upon the signal content of the input test chart). Note that in this thesis, target-dependency is referred to by the related, broader term of scene-dependency. Linearising the signal using look-up tables (LUT) before computing the MTF reduces errors resulting specifically from non-linear system tone reproduction [23, p. 6], [160] but does not account for the mentioned local spatial scene-dependency.

Moreover, since edges, sinusoids and noise signals have limited relation to the average pictorial scene signal, their interaction with non-linear ISP algorithms means that the abovementioned measurements consistently under-estimate or over-estimate the average real-world signal transfer of the system (i.e. they are biased); $\bar{F}(s)$ in Figure 1.4(b) defines the average real-world signal transfer of such systems. The mentioned measurement methods using edges, sinusoids and noise signals also do not characterise perceived texture loss effectively [17]. The latter is a primary driver of overall perceived quality in today's non-linear capture systems [3]. These limitations motivate the following suggestion: *if the MTF aims to describe the real-world signal transfer performance of a system that implements non-linear content-aware processing, then it should be derived either from a test chart that represents the signal properties of pictorial scenes, or better yet, from images of scenes*. This suggestion only applies, however, if deriving the MTF in such a way does not result in significantly higher levels of measurement error than current MTF measurement methods.

It is now common for the dead leaves test chart [16], [17] (shown in Figure 2.10(b)) to be employed when characterising non-linear capture system signal transfer. This test chart

relates more closely to natural scene signals than edges, sinusoidal signals and noise. It simulates natural scene textures using a stochastic model. Circles of random density are overlaid, reproducing occlusion phenomena and varying contrast levels. It also replicates the average power spectrum of natural scenes, among other natural scene statistics (NSS) [17]. Other useful properties are its shift, scale and rotational invariance. This test chart is employed when measuring the MTF with the intention of triggering non-linear image processing at ‘natural’ levels (i.e. the same level as when the system processes an “average pictorial scene”). The resultant MTFs are more indicative of the average real-world performance of non-linear systems than measures derived from edges, sinusoidal signals and noise. They also characterise more effectively the texture loss in such systems. There are three dead leaves MTF implementations, summarised below.

The *original dead leaves measurement implementation* [17] employs Equation 2.29 to calculate the MTF. Since system noise is unaccounted for, $PS(u)_{output}$ increases erroneously as the power of the noise increases, biasing the MTF.

The *direct dead leaves measurement implementation* [19] is defined in Equation 2.31. It attempts to remove the mentioned bias by subtracting the NPS of the system, $NPS_{output}(u)$, from the output power spectrum, $PS_{output}(u)$. $PS_{input}(u)$ is the input power spectrum, u is spatial frequency, and the rotational average of $PS(u, v)$ in Equation 2.30 yields $PS(u)$. The NPS is measured using a uniform-tone patch with limitations discussed in Section 2.3.3. Error in the NPS is carried into the MTF. This dead leaves measurement implementation is employed by the texture blur metric of the IEEE P1858 CPIQ standard [22, p. 53].

$$MTF(u) = \sqrt{\frac{PS_{output}(u) - NPS_{output}(u)}{PS_{input}(u)}} \quad (2.31)$$

The recently standardised *intrinsic dead leaves measurement implementation* [20] compares signal transfer with respect to the cross-spectrum of dead leaves signals, which contains both amplitude and phase information. This measurement implementation is capable of distinguishing between real image structures and certain image processing artefacts and delivers consistent measurements for systems that generate significant noise. It characterises the performance of the lens and imager only and is virtually unaffected by reversible (linear) processing such as contrast stretching or sharpening, provided that clipping or information loss does not occur [161]. The intrinsic implementation is less relevant than the direct

implementation to the aims of this thesis. This is because reversible image processing affects the image quality and perceived sharpness of many capture systems (including non-linear systems that implement a combination of linear and non-linear ISP algorithms). Thus, any image quality metric that aims to account for system scene-dependency should account for the effect of both reversible and non-linear image processing on system signal transfer. The intrinsic implementation is not employed by the IEEE P1858 CPIQ standard [22].

The various dead leaves implementations provide convenient approximations of the average real-world signal transfer characteristics of systems implementing non-linear content-aware processing. No prior art has investigated, however, whether the dead leaves test chart triggers non-linear content-aware processing at the same levels as the “average natural scene” and thus derives the average real-world MTF of such systems accurately and appropriately. This assumption may not be achieved in practice since the mathematically-generated dead leaves chart has limited relation to the diverse and complex spatial signals in real pictorial scenes. Further, no MTF measured from a single printed test chart can account fully for scene-dependent variations in the signal transfer of non-linear systems. This would require MTFs to be measured with respect to signals from images of real scenes, as suggested, to trigger non-linear content-aware processing appropriately. Branca *et al.* [7] were the first to achieve this by revising the dead leaves method. Their methodology is defined, evaluated critically, and extended by the scene-and-process-dependent MTF (SPD-MTF) framework that is proposed in Chapter 5 of this thesis.

2.3.3 Noise

Noise is defined, objectively, as random or non-random spurious image information [1, p. 346], [60]. It is perceived as a particular type of texture [16] and produces sensations of the corresponding subjective attribute of noisiness, which is detrimental to image quality generally [1, p. 433], [46, p. 220].

Scene susceptibility to noisiness relates to the amount of scene texture [58], among other factors, as indicated by Figure 2.5. The perceived impact of physical noise on image quality is also scene-dependent, where higher quality images are most affected [111], [112], [137]. Adding limited levels of noise to images of scenes can increase their perceived sharpness under some conditions [162], [163]. This is either because observers confuse noise with

memory representations of real scene textures [163], or they associate the spurious high frequencies with fine detail only present in images from sharper, higher-quality systems [33].

Noise exists in the output of all imaging system components and processes. Sources of physical noise at all stages of the image capture, processing and display chain are described below in the approximate order that they appear. Visual noise is also experienced by the observer, dependent upon the image presented and viewing conditions. Visual noise models are presented in Section 2.5.

For an ideal detector, the output noise is limited to noise in the input signal. For real detectors, noise is added to the input signal that already contains noise. Noise in the input signal is referred to as Poisson exposure noise, photon noise, or shot noise. It is caused by discrete local variations in flux in the radiant intensity image due to the quantum nature of light [2, p. 48] (Figure 2.11(a)). The number of photons arriving at the active area of each pixel on the sensor is random and is modelled by the Poisson distribution. The SNR of Poisson noise relates to the number of exposure quanta, q [1, p. 442], [6, p. 153], [164, p. 278]; thus, even noise in a hypothetical ideal linear system is scene-dependent. When q is high the SNR increases, and Poisson noise is less perceptible. At low light levels, however, it is common for Poisson noise to dominate over most sources of noise added by the system. The exception is read noise, which can range from 3 electrons to 8 electrons at low and high sensor temperatures, respectively, significantly affecting quality at low light levels [165] (Figure 2.11(b)).

$$SNR = \frac{q}{\sqrt{q}} = \sqrt{q} \quad (2.32)$$

Photoelectric noise and electronic noise are introduced when the radiant intensity image is converted to an electrical signal and processed by capture system components, respectively. Jenkin [166, pp. 433, 442] describes the following sources of this noise. Fixed pattern noise (FPN) is consistent between frames and is caused by imperfections in the CFA elements, including dark/dead pixels that are insensitive to light, hot pixels that saturate too quickly, or dust. Dark signal non-uniformity (DSNU), or dark current noise, is introduced when the number of thermally-generated electrons varies from pixel to pixel (Figure 2.11(d)). It relates to the exposure duration and temperature of the sensor. It exists in the output signal irrespective of whether there is any input signal and can be mitigated by cooling. Photoresponse non-uniformity (PRNU) results from minor differences in pixel sensitivity

due to variations in their pixel well size/area or the properties of the substrate (Figure 2.11(c)). Quantisation noise is introduced when the continuous electrical signal is binned into discrete pixel values. Amplification noise consists of white noise and flicker noise. Both are dependent on the sampling rate. Further, reset noise results from slight variations in the charge stored in each photoelement after they have been reset to the reference voltage after each exposure. Onboard capture system ISP and further imaging chain components and processes alter the characteristics of noise from all the above sources and can add further noise to the displayed image [19]. The ISP algorithms with the greatest effect upon noise include demosaicing, denoising, sharpening, tone-mapping and compression.

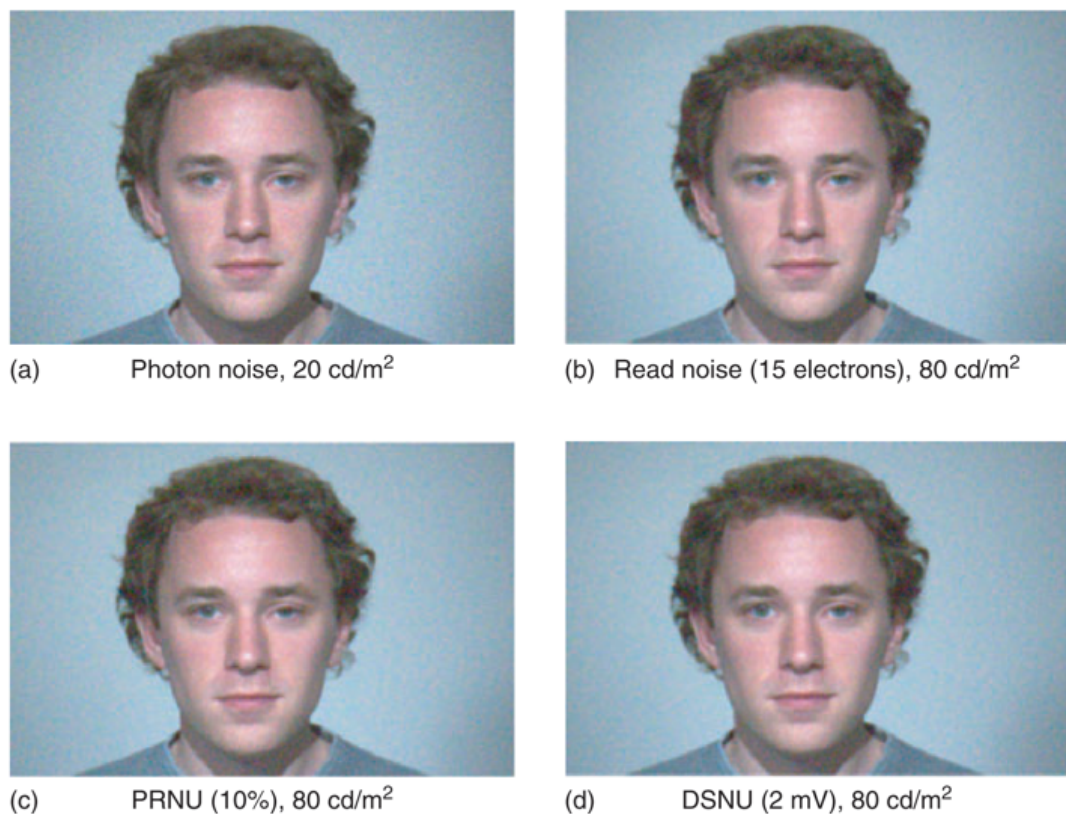


Figure 2.11 Simulations by Farrell & Wandell [167] of: (a) temporal photon noise, and (b) read noise, as well as fixed-pattern noise caused by (c) photoresponse non-uniformity (PRNU) and (d) dark signal non-uniformity (DSNU) in image capture systems.

The various objective measures of image noise will now be defined and discussed. The standard deviation, or RMS noise, is the most common image noise measure but accounts for the amplitude of spurious image information only [41, p. 28]. A uniform neutral grey patch is first captured under uniform illumination from an appropriate target such as Figure 2.12. The RMS noise is then calculated using Equation 2.3 where L_{ij} is the

luminance/reflectance of each pixel in the region ($M \times N$) and \bar{L} is the sample mean. Imaging system noise is characterised at different exposures by plotting RMS noise measurements from patches of various luminances vs the patch luminance.



Figure 2.12 Image Engineering™ TE264 test chart [130] for noise measurements according to ISO 15739 [168]. The chart contains 20 uniform-tone patches.

The RMS noise measure describes stochastic noise adequately, which has a grainy appearance comparable to Figure 2.12(a). However, the combined effect of the various sources of capture system noise and non-linear content-aware ISP algorithms such as denoising or sharpening often results in noise that is visibly textured [41, p. 32]. Such noise can relate to characteristics of the input signal as indicated by Figure 2.3. Figure 2.13 shows that the RMS noise measure is incapable of accounting for such spatial variations.

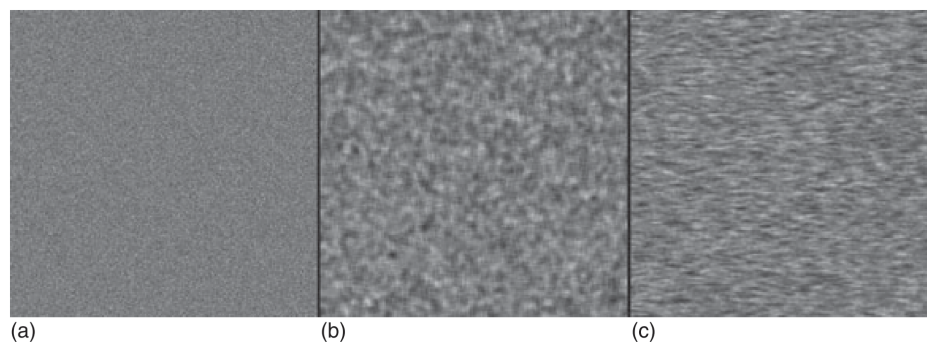


Figure 2.13 Noise with identical mean signal and standard deviation (i.e. the same level of root mean square (RMS) noise) but varying correlation [41, p. 32]: (a) low two-dimensional (2D) covariance, (b) high 2D covariance, (c) low vertical covariance and high horizontal covariance.

Measures that account for both the spatial characteristics and amplitude of noise are the Autocovariance Function (ACF), Autocorrelation Function, and the NPS (or Wiener

spectrum). The ACF describes the spatial relationship (covariance) between the intensity of each pixel in terms of their separation [41, p. 31]. Noise with a grainy appearance has lower covariance than noise with a patterned structure, as demonstrated by Figure 2.13. Burns [41, p. 31] defines the ACF for digital systems below. d_{ij} is the luminance/reflectance of each pixel in the region $(M \times N)$ of a captured uniform patch. m_d is the sample mean.

$$A_{k,l} = \frac{1}{(MN - 1)} \sum_{i=1}^{M-k} \sum_{j=1}^{N-l} (d_{i,j} - m_d)(d_{i+k,j+l} - m_d) \quad (2.33)$$

Dividing the ACF, $A_{k,l}$, by the variance, σ^2 , yields the autocorrelation function of the system, $C'_{k,l}$, that is equal to 1 at the origin [41, p. 31].

$$C'_{k,l} = \frac{A_{k,l}}{\sigma^2} \quad (2.34)$$

When the requirements of linear system theory are satisfied (Appendix A), Dainty and Shaw [6, p. 222] state that convolving the original (input) autocorrelation function, $C_{k,l}$, twice with the PSF, $h(x, y)$, yields the measured (output) autocorrelation function, $C'_{k,l}$.

$$C'_{k,l} = C_{k,l} \circledast h(-x, -y) \circledast h(x, y) \quad (2.35)$$

The NPS is the standard measure of imaging system noise [2, p. 2], [158]. It defines the power of noise with respect to spatial frequency. It holds its foundation in linear system and communications theory [6, Sec. 6.3] and thus requires Equations A1 to A4 to be satisfied. It is implemented as a noise parameter by many multivariate IQMs for capturing systems [12], [169] that cascade it with MTF and CSF measures. Welch [41, p. 34], [170] defines the NPS of a digital system below. $\langle \dots \rangle_B$ is the average over B blocks, $I(x, y)$ is the luminance at pixel coordinates (x, y) in the region $(M \times N)$ of a captured uniform patch. u and v are spatial frequencies. Δx and Δy are sampling intervals with respect to x and y .

$$NPS(u, v) = \frac{1}{NM\Delta x\Delta y} \left\langle \left| \sum_{x=M/2+1}^{M/2} \sum_{y=N/2+1}^{N/2} I(x, y) e^{-2\pi i(ux+vy)} \right|^2 \right\rangle_B \quad (2.36)$$

The NPSs, $NPS(u, v)$, of two components or processes denoted by the subscripts A and B are combined additively:

$$NPS_{A+B}(u, v) = NPS_A(u, v) + NPS_B(u, v) \quad (2.37)$$

The NPS and autocorrelation function are Fourier transform pairs according to the Wiener-Khintchin theorem [6, p. 222]. Thus, their relationship as image noise measures is comparable to the relationship between the MTF and PSF measures of signal transfer [6, p. 222]. If the Fourier transform of both sides of Equation 2.35 is taken, the measured (output) NPS, $NPS'(u, v)$, is shown to be equal to the original (input) NPS, $NPS(u, v)$, multiplied by the squared modulus of the OTF, $OTF(u, v)$ [6, p. 223], or the square of the MTF, $M(u, v)$. This relationship also applies to signal power spectra and is fundamental to MTF measurements from white noise and the dead leaves test chart, defined in Equations 2.29 and 2.31, respectively.

$$NPS'(u, v) = NPS(u, v)|OTF(u, v)|^2 = NPS(u, v)M(u, v)^2 \quad (2.38)$$

Each of the mentioned noise measures is derived generally using noise images, $I(x, y)$, from imaged uniform-tone patches, where $g(x, y)$ is the intensity of the captured patch, Equation 2.39. Obtaining noise images in this way is convenient, since the expected value of the captured patch, $\bar{g}(x, y)$, is approximately constant at all coordinates, provided that the patch is illuminated uniformly and lens shading correction is applied (or the patch is captured in sensor areas where lens shading is minimal).

$$I(x, y) = g(x, y) - \bar{g}(x, y) \quad (2.39)$$

Many IQMs employ noise measurements derived in this way. These metrics attempt to describe the perceived performance level of the system either with respect to a particular input scene or the “average pictorial scene” (i.e. concerning the average real-world performance of the system). For such metrics (and the measurements they are derived from) to be appropriate, it must be assumed that the noise in the captured patch represents the characteristics of noise introduced either to the scene in question, or the “average scene”. This assumption is justified for linear systems, for which, even considering the slight scene-dependency in Poisson noise and other system noise, there is (within reason) an NPS that fully specifies the system. However, the assumption fails for systems implementing non-linear content-aware denoising and sharpening. For these systems, the structure and magnitude of noise is scene-dependent (and target-dependent) due to interactions between input scene (and test chart) content and the ISP algorithms (refer to Section 2.2.1 for more

information). Since uniform patches provide ideal input conditions for non-linear content-aware denoising algorithms, they are significantly less noisy than real scenes when captured by such systems, as indicated by Figure 2.3(d). Noise measures derived from uniform patch signals, therefore, underestimate the average real-world noise of such systems, as well as noise in a given captured scene. These inaccuracies are carried into IQM measurements and are compounded if non-linear content-aware sharpening is also applied (Section 2.2.1).

Consequently, as with the MTF, it is suggested that more suitable NPS measurements should either be derived directly from images of pictorial scenes, or from test charts that replicate the signal properties of scenes, provided this does not significantly increase measurement error. In this thesis, the scene-and-process-dependent NPS (SPD-NPS) framework, proposed in Chapter 4, is used to deliver such measurements. The various SPD-NPS measures proposed in this thesis are either measured from images of pictorial scenes, or dead leaves signals that model the “average scene spectrum”. Thus, they account more appropriately than the uniform patch NPS for the effect of non-linear content-aware ISP algorithms.

2.4 Signal-to-Noise Measures

The relative intensity of image signals versus noise affects the capability of the HVS to detect, discriminate and understand the content of a given image. Thus, objective and subjective image quality has been described as a signal-to-noise-based problem for decades. In 1948, Shannon [171] founded the field of information theory. This defined the general signal-to-noise relationship demonstrated in numerous applications of communication as the Channel Capacity, C , given by Equation 2.40 [171, p. 47]. When applied to imaging systems, S and N are the signal and noise energy and W is the channel bandwidth, corresponding to sharpness [172].

$$C = W \log_2 \left(\frac{S}{N} + 1 \right) \quad (2.40)$$

Schade [43] applied Shannon’s approach to imaging systems development, resulting in Equation 2.41 [173, p. 631] that expresses the photographic Information Capacity, c , where $S(u_x, u_y)$ and $N(u_x, u_y)$ are 2D signal and noise power spectra, respectively, which can be computed using Equations 3.5 and 3.6, respectively for analog systems; u_x and u_y are spatial frequency. This equation is at the foundation of Töpfer and Jacobson’s *Perceived*

Information Capacity (PIC) [169] image quality metric that is defined in Equation 3.3.

$$c = \frac{1}{2} \int_0^\infty \int_0^\infty \log_2 \left(\frac{S(u_x, u_y)}{N(u_x, u_y)} + 1 \right) du_x du_y \quad (2.41)$$

Higgins [136] expresses imaging system information capacity, c , in a simpler form, below, which is core to the *Effective Pictorial Information Capacity (EPIC)* metric that is defined in Section 3.1.1. The number of independent levels capable of being reproduced at each pixel and the number of pixels in the image are represented by m and n , respectively.

$$c = n \log_2 m \quad (2.42)$$

Simple, global SNR measures also correlate to some extent with perceived noise intensity and image quality. The SNR can be expressed as the reciprocal of the coefficient of variation, as shown in Equation 2.43, where μ_{sig} is the expected value or mean of the signal and σ_{bg} is the standard deviation of the noise or background. However, the signal transfer characteristics of the system and the structure of the noise are unaccounted for and often vary with spatial frequency.

$$SNR = \frac{\mu_{sig}}{\sigma_{bg}} \quad (2.43)$$

The Detective Quantum Efficiency (DQE) is a more comprehensive signal-to-noise-based performance measure for capture systems. It is essentially a measure of “useful” quantum efficiency and describes how effectively the system reproduces information from a stream of quanta (photons), relative to an ideal photon-counting detector [6, p. 28]. The DQE is defined below. $SNR_{out}(u)$ is the SNR of the system, $SNR_{in}(u)$ is the input SNR (i.e. the SNR of an ideal detector [6, p. 153]) and u is spatial frequency. The DQE accounts for the MTF, NPS and tone transfer characteristics of a given system [174], thus providing all necessary information to describe its spatial image quality [175]. Systems with high DQE image small low-contrast objects effectively.

$$DQE(u) = \frac{SNR_{out}^2(u)}{SNR_{in}^2(u)} \quad (2.44)$$

For an ideal detector, the lower limit of noise is the quantum fluctuations [6, p. 152]. Equation 2.32 shows the SNR of Poisson noise is equal to the square root of the number of exposure quanta. Thus, $SNR_{in}^2(u)$ is equal to the number of quanta, q_A [6, p. 153].

$$SNR_{in}^2(u) = q_A \quad (2.45)$$

Although the DQE is normalised conveniently to unity, Keelan [176] states that the NEQ is a more appropriate performance indicator for capture systems. The NEQ relates to the DQE, below [177]. q_A is the number of exposure quanta.

$$DQE(u) = \frac{NEQ(u)}{q_A} \quad (2.46)$$

Thus, the NEQ is effectively the output SNR squared, with respect to spatial frequency, u .

$$NEQ(u) = DQE(u) \cdot SNR_{in}^2(u) = SNR_{out}^2(u) \quad (2.47)$$

Burns [177] defines the 1D NEQ below. $MTF(u)$ is the MTF, $NPS(u)$ the NPS and q_A the number of quanta per unit area. G_d is the mean level transfer gain, dq_A/do , where o is the output signal.

$$NEQ(u) = \frac{q_A^2 G_d^2 MTF^2(u)}{NPS(u)} \quad (2.48)$$

If the MTF and NPS of a given system are input referenced and computed in a linear space, G_d can be assumed to be 1. Thus Keelan [176], [178] defines the 2D NEQ below, where μ is the mean linear signal. The 1D NEQ can be computed either by substituting the 1D MTF and NPS into this equation or as the rotational average of $NEQ(u, v)$.

$$NEQ(u, v) = \frac{MTF^2(u, v)}{NPS(u, v)/\mu_A^2} \quad (2.49)$$

The NEQ describes the physical number of quanta counted by the system, q'_A , after accounting for noise, considering an ideal detector is capable of counting q_A quanta at the same exposure level [6, p. 156]. Therefore, it provides a more relevant comparison of systems at different exposure levels than the DQE [176]. The following investigations have

employed the NEQ in capture system sensor [176]–[178] and image quality modelling [178] successfully.

The NEQ has also long been used in signal detection theory, including in theoretical models of the ideal observer [179], [180] in simple idealised scenarios where the signal is known exactly (SKE) and the background is known exactly (BKE). One such model, the detectability index, d' , is defined below for a signal with mean-normalised frequency spectrum, $S(u_x, u_y)$, in terms of the two-dimensional (2D) NEQ of the system, $NEQ(u_x, u_y)$ [178], [181]. It has proven valuable in the design and optimisation of medical imaging equipment [181]. Models for more complex theoretical observers have been developed upon it. These models either account for correlated noise or are applicable in situations where target and background spectra are not known exactly [181].

$$d'^2 = \int NEQ(u_x, u_y) S^2(u_x, u_y) du_x du_y \quad (2.50)$$

The NEQ, DQE, and photographic information capacity all are founded upon linear system theory. Consequently, when the requirements of Appendix A are not fulfilled, inaccuracies in the traditional MTF and NPS measures (discussed in Sections 2.3.2 and 2.3.3, respectively) are carried into each of these signal-to-noise measures.

2.5 Visual Models

The visual models employed by IQMs can be classified broadly as top-down or bottom-up approaches. Top-down approaches are hypothetical and can be characterised as black-boxes or as the combination of a set of black-box sub-processes. They often attempt to describe higher-level processes concerning image quality judgement as well as low-level vision. They are implemented in the *Computational IQM (CP-IQM)* genre that is of less interest to this project for reasons given in Section 3.1.4, where this genre of IQMs is also defined. The same section discusses critically the top-down visual models employed by the most relevant CP-IQM approaches to this thesis.

Bottom-up approaches are employed by all other IQM genres reviewed in this thesis. They model low-level visual processes in a logical and mechanistic step-by-step format with direct relation to psychophysical measurements and the neurophysiology of the HVS. The most commonly employed bottom-up spatial HVS model is the CSF, defined later in this section.

It is employed as a weighting function by most spatial IQMs for capturing systems to account for the perceptual significance of MTF and NPS system performance measurements.

A number of multivariate IQMs for capturing systems also employ bottom-up modelling of visual noise. Noteworthy examples [12], [169] implement Barten's [182] visual noise model, defined below. The IQM from the IEEE P1858 CPIQ standard [22, p. 72] implements a perceptually calibrated *visual noise attribute metric* for image quality loss due to noise. It is defined in Section 3.1.2 and also applies the CSF.

Visual noise is dependent on the viewing distance and the luminance and size of the presented image [182]. Barten's [182] model for the total power of internal noise in the eye, $NPS_{Visual}(u)$, vs spatial frequency in cycles/degree, u , is given below. ϕ_{Ph} and ϕ_{Neu} are the visual photon noise and neural noise, given by Equations 2.52 and 2.53, respectively.

$$NPS_{Visual}(u) = \phi_{Ph} + \phi_{Neu}(u) \quad (2.51)$$

Visual photon noise, ϕ_{Ph} , is modelled as follows [182] where L is luminance, d is the pupil diameter of the eye in mm given by Equation 2.57. η is the quantum efficiency of the eye, recorded as 21% by Ford [9, p. 28] in a comparable study. ρ is the photon conversion factor of the eye, recorded as 1.285×10^6 photons/(s Troland degree²) for Illuminant A under photopic viewing conditions [183, p. 61]. T is the integration time of the eye given by Töpfer and Jacobson [169] as 0.1s for a stationary signal under such conditions.

$$\phi_{Ph} = \frac{1}{\eta \rho \frac{\pi}{4} d^2 L T} \quad (2.52)$$

Neural noise, ϕ_{Neu} , is modelled below [182] where $MTF_{Lat}(u)$ is the MTF of the lateral inhibition process [183, p. 33], u is spatial frequency, and u_0 is the threshold at which lateral inhibition ceases. ϕ_0 is the intensity of neural noise. Töpfer and Jacobson [169] state u_0 is 11 cycles/degree and ϕ_0 is 1.5×10^{-7} degrees² for a typical integration time of 0.1s.

$$\phi_{Neu}(u) = \frac{\phi_0}{MTF_{Lat}(u)} = \frac{\phi_0}{1 - e^{-u^2/u_0^2}} \quad (2.53)$$

The CSF describes the sensitivity of the HVS to luminance (or chrominance) contrast in terms of spatial frequency. It varies with respect to the signal's luminance, spatial frequency [183, Ch. 3], orientation [184], [185], phase coherency [184], and the intensity of masking

noise [183, Ch. 6]. From an imaging systems development perspective, the most important relationship is between luminance contrast and spatial frequency, with consideration of noise and other masking signals.

Contrast sensitivity is measured conventionally with respect to simple, narrow-band stimuli, such as sine-waves, edges or noise patches. Traditionally, the CSF describes the detection of these stimuli in isolation. Such measurements are obtained by varying the contrast of the stimulus until a threshold of one just-noticeable difference (JND) is reached between the (test) stimulus and an identical luminance patch of zero contrast (the reference stimulus).

To quantify one JND between the stimuli, observers are shown the test and reference stimulus on a calibrated display under strict viewing conditions. The observers are asked to indicate which the test stimulus is. Their votes are divided 50/50 between the test and reference stimulus when differences between the stimuli are imperceptible. For the difference between the stimuli to be exactly one JND, the observers identify the test stimulus correctly 50% of the time (i.e. their votes are split 75/25 in favour of the test stimulus).

Contrast signals above and below the threshold of one JND are termed suprathreshold and sub-threshold, respectively. The traditional CSF, or contrast detection CSF, is given by the reciprocal of this threshold vs spatial frequency (Figure 2.14, broken line). It peaks at around 1-4 cycles/degree for photopic vision.

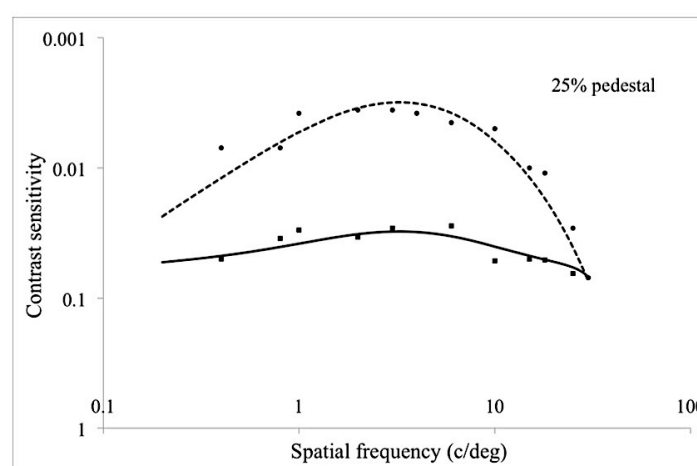


Figure 2.14 Barten's detection (broken line) [186] and discrimination (solid line) [183, p. 143] contrast sensitivity functions (CSF) [187].

There are a number of models used to predict the CSF. The simplest and most popular are black-box models for contrast detection that account for spatial frequency only. For example,

Johnson and Fairchild's luminance and chrominance CSFs [188] (defined by Equations 2.54 and 2.55, respectively) are used by the IEEE P1858 CPIQ standard [22, p. 72]. The luminance CSF employs Movshon's three-parameter exponential model [189], with different constants given in Table 2.1. u is spatial frequency in cycles/degree.

$$CSF_{luminance}(u) = \frac{a_1 \times u^{c_1} \times e^{-b_1 \times u}}{K} \quad (2.54)$$

$$CSF_{chrominance}(u) = \frac{a_1 \times e^{-b_1 \times u^{c_1}} + a_2 \times e^{-b_2 \times u^{c_2}} - S}{K} \quad (2.55)$$

Parameter	Luminance CSF (A)	Red-Green Chrominance CSF (C ₁)	Blue-Yellow Chrominance CSF (C ₂)
a1	75	109.1413	7.0328
b1	0.2	0.0004	0
c1	0.8	3.4244	4.2582
a2		93.5971	40.691
b2		0.0037	0.1039
c2		2.1677	1.6487
K	75	202.7384	40.691
S		0	7.0328

Table 2.1 Input parameters for Johnson and Fairchild's luminance and chrominance CSFs [188] as implemented in the IEEE P1858 Camera Phone Image Quality (CPIQ) standard [22, p. 72].

Barten's contrast detection model [186] is given below. It accounts for the display's angular size in degrees, w , its luminance, L , and the stimulus' spatial frequency in cycles/degree, u .

$$CSF_{Barten}(u) = \frac{1}{M_t(u)} = au \cdot e^{(-bu) \cdot \sqrt{1 + c \cdot e^{(bu)}}}, \quad (2.56)$$

where

$$a = \frac{540(1 + 0.7/L)^{-0.2}}{1 + \frac{12}{w(1 + u/3)^2}}, \quad b = 0.3(1 + 100/L)^{0.15}, \quad \text{and} \quad c = 0.06$$

Barten's mechanistic detection model [183, p. 39] is defined by Equation 2.57 and expands upon the simpler model above. It employs relevant physiological and physical variables, thus bridging the gap between contrast detection/discrimination and the underlying neurophysiology. $M_{opt}(u)$ is the optical MTF of the eye, k the SNR, T the integration time of the eye. X_o and Y_o are horizontal and vertical angular stimulus sizes, respectively, and X_{max} is the maximum integration area, all given in degrees. N_{max} is the maximum number

of integration cycles, η the quantum efficiency of the eye, E the retinal illuminance in Troland and p is the photon conversion factor. Φ_0 is the spectral density of neural noise, u is spatial frequency in cycles/degree, u_0 is the frequency where lateral inhibition ceases. L is the average luminance of the stimulus, σ the standard deviation of the LSF of the eye and d is the diameter of the pupil in mm. The constants σ_0 and C_{ab} are approximately 0.5 and 0.08 arc min/mm, respectively.

$$CSF_{Barten}(u) = \frac{1}{M_t(u)} = \frac{M_{opt}(u)}{k \sqrt{\frac{2}{T} \left(\frac{1}{X_o Y_o} + \frac{1}{X_{max}^2} + \frac{u^2}{N_{max}^2} \right) \left(\frac{1}{\eta p E} + \frac{\Phi_0}{1 - e^{-\left(\frac{u}{u_0}\right)^2}} \right)}} \quad (2.57)$$

where

$$M_{opt}(u) = e^{-2\pi^2 \sigma^2 u^2}, \quad \sigma = \sqrt{\sigma_0^2 + (C_{ab} d)^2}, \quad d = 5 - 3 \tanh\left(0.4 \log\left(\frac{L X_o Y_o}{40^2}\right)\right)$$

Contrast discrimination describes sensitivity to variations in suprathreshold signal contrast. Discrimination sensitivity is generally lower and flatter than detection sensitivity (Figure 2.14) due to the effects of visual masking [183, Ch. 7]. For isolated narrow-band stimuli, this is attributed to contrast constancy [190] where neural processing “evens out” suprathreshold contrast perception with respect to spatial frequency [113]. This provides a useful de-blurring effect that compensates for the physical limitations of the HVS. Peli [123] argues, however, that contrast constancy is less applicable to contrast discrimination in complex broadband images because neural processes are activated simultaneously by overlapping contrast signals of varying frequency and orientation.

Barten’s contrast discrimination model [183, p. 143], $CSF_{Discrim}(u)$, is based upon measurements of dipper functions. k is the Crozier factor [191], [192] that is approximately equal to 3.0. It describes the minimum signal-to-noise level for signal detection [187]. $m_t(u)$ is the modulation threshold required for contrast detection, given by Equation 2.56 or 2.57. $m_0(u)$ is the modulation of the reference signal, u is spatial frequency in cycles/degree.

$$CSF_{Discrim}(u) = \left[\sqrt{\frac{m_t^2(u) + 0.04 k_{crozier}^2 m_0^2(u)}{1 + 0.004 k_{crozier} \frac{m_0(u)}{m_t(u)}}} + m_0^2(u) - m_0(u) \right]^{-1} \quad (2.58)$$

Triantaphillidou *et al.* [187] have extended Barten's models to describe the detection and discrimination of RMS contrast in single-octave spatial frequency bands from images of pictorial scenes. Each band was extracted in the spatial frequency domain using Peli's log cosine filters [118]. The *Isolated CSF (iCSF)* [35] and *Isolated Visual Perception Function (iVPF)* [187] model the detection and discrimination of contrast in each band in isolation, respectively. The iCSF is described appropriately at each band's spatial frequency by either of Barten's detection models [183, p. 36], [186] (i.e. Equations 2.56 or 2.57). Likewise, Barten's discrimination model [183, p. 143] (Equation 2.58) expresses the iVPF [187] when m_0 is the pedestal contrast (i.e. the RMS contrast of the band) and m_t is given by Equation 2.56 or 2.57.

The *Contextual CSF (cCSF)* [35] and *Contextual Visual Perception Function (cVPF)* [8] model detection and discrimination, respectively, of the RMS contrast of a band when all other bands are at full contrast. Both functions are shown in Figure 2.15.

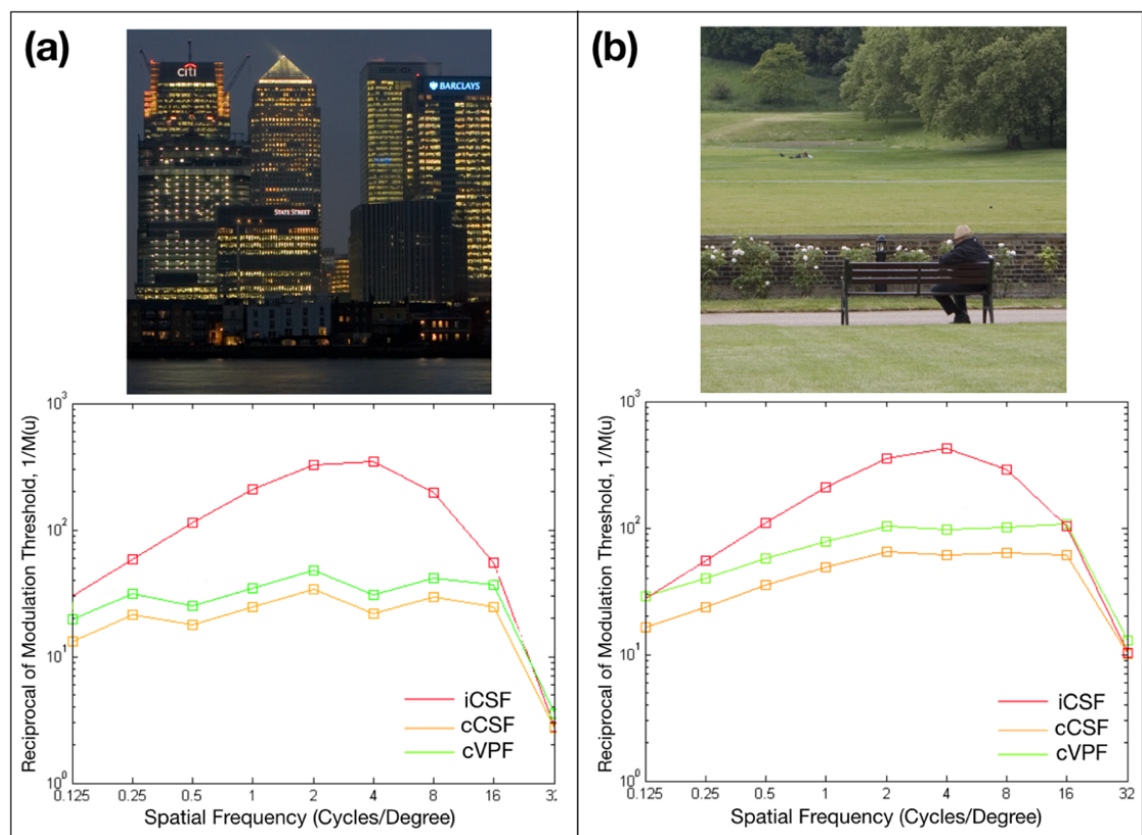


Figure 2.15 Isolated contrast detection (iCSF) [35] and contextual contrast detection (cCSF) [35] and discrimination (cVPF) functions [8] for the 'Buildings' (a) and 'Bench' (b) images shown in Appendix C. Adapted from [14].

To account for visual masking, the cCSF [35] (given by $cCSF(u)$ in Equation 2.59) models as noise the contrast signals in “flanking bands” adjacent to the band in question [8], using the Linear Amplification Model (LAM) [193]. K is a scene-dependent constant, $c_s(u)$ is the RMS contrast spectrum of the image (defined in Section 2.3.1), $c_i(u)$ is the contrast threshold for isolated detection of each band (i.e. the reciprocal of the iCSF) and u is spatial frequency in cycles/degree.

$$cCSF(u) = [Kc_s(u)^2 + c_i(u)^2]^{-0.5} \quad (2.59)$$

Substituting the cCSF (Equation 2.59) into Barten’s discrimination model [183, p. 143] (Equation 2.58) yields the cVPF that is defined by Equation 2.60 [125]. $c_c(u)$ is the contrast threshold required to achieve contextual detection for each band, i.e. the reciprocal of the cCSF, $c_s(u)$ is the RMS contrast spectrum, $k_{crozier}$ is the Crozier factor of approximately 3.0, u is spatial frequency in cycles/degree.

$$cVPF(u) = \left[\sqrt{\frac{c_c^2(u) + 0.04k_{crozier}^2 c_s^2(u)}{1 + 0.004k_{crozier} \frac{c_s(u)}{c_c(u)}}} + c_s^2(u) - c_s(u) \right]^{-1} \quad (2.60)$$

There is ongoing debate regarding whether image quality is a function of suprathreshold vision [36], [194] or threshold vision [185], or a combination of both [8], [36], [41, p. 43], [183, p. 4]. Most IQMs based upon linear system theory either cascade system performance measures with threshold CSFs or weight image differences with them [12], [169], [186], [195]–[199]. This has been shown to account for the perceptual relevance of these measurements under the viewing conditions [14]. However, it assumes a direct relationship exists between threshold vision and image quality/fidelity perception. Although there lacks a better alternative, Rogowitz *et al.* [200] have criticised the implementation of threshold CSFs in IQMs, suggesting that image quality judgement involves the subjective scaling and comparison of suprathreshold image content related to quality attributes and artefacts. Likewise, Haun and Peli [113] stress the importance of contrast thresholds when evaluating image quality and have measured the perceived contrast magnitude of spatial frequency bands from pictorial scene images [123], [124]. Their measurements display a similar band-pass profile to traditional CSFs and peak at similar frequencies [123], [124], but they have not been modelled and are subject to inter-observer inconsistency.

During image quality/fidelity evaluation, attributes and artefacts are detected and discriminated within the context of other suprathreshold image signals. Consequently, visual masking models are employed alongside detection/discrimination CSFs in many metrics that aim to describe fidelity [34] (these metrics are defined in Section 3.1.3); it is expected that such modelling would also benefit spatial capture system IQMs. The cCSF and cVPF account for visual masking with respect to complex suprathreshold signals and are expected to be more relevant functions for image quality modelling than both traditional CSFs, the iCSF and the iVPF. The cCSF and cVPF are both scene-dependent with respect to the contrast spectrum of the image and, therefore, may account for scene-dependent gain and offset in correlations between IQM scores and observer image quality ratings (as shown in Figure 1.5). Further, the shape of the cCSF and cVPF are often similar to one another (Figure 2.15). This suggests that contextual contrast detection and discrimination processes may collapse into a single scene-dependent function for contextual human vision, which, if true, would be advantageous for the development of adaptive, scene-dependent IQMs.

Since image quality evaluation involves comparison with internal memory representations [13], [55], it is unlikely that all relevant visuo-cognitive activity is accounted for by traditional CSFs, cCSFs, cVPFs, or any other bottom-up models of lower-level vision [14]. However, the cCSF/cVPF are expected to provide an appropriate half-way point after lower-level visual processing, upon which a hypothetical ideal mechanistic IQM could expand.

The author has developed *optimal contrast weighting (OCW)* [14] functions that describe relationships between CSFs/cCSFs/cVPFs and optimised spatial image quality, to investigate separately the higher-level cognitive processes of image quality judgement. Contrast optimisation was carried out by weighting nine single-octave contrast bands, ranging from 0.125 to 32 cycles/degree. This research found the cCSF and cVPF were more appropriate base functions for OCWs [14] than traditional CSFs that had been tested previously [138], [139]. Further, OCWs formed from high-pass filtered cCSFs and cVPFs outperformed Adobe™ Photoshop's sharpening filters [14]. Although image quality optimisation and image quality modelling are not identical, the successful application of cCSFs/cVPFs in OCWs suggests they should also improve IQMs. Further research carried out in laboratories at the University of Westminster [201] has demonstrated that band adjustments are capable of compensating for image quality losses due to visual impairment.

2.6 Summary

The reproduction of spatial luminance contrast signals influences subjective image quality significantly, particularly the attributes of resolution, sharpness, noise and contrast. Spatial imaging performance characterisation is fundamental to capture system design and quality modelling. The MTF, NPS and NEQ are used to characterise system signal transfer, noise and signal-to-noise performance, respectively. They are summarised in Table 2.2 and are founded on linear system theory. Current implementations of these measures are derived using test charts. They often characterise scene-dependent systems applying non-linear ISP algorithms inaccurately. This affects the accuracy of relevant spatial IQMs. This thesis develops novel scene-and-process-dependent system performance measures and IQMs to address these limitations.

Performance Measure	Appropriate Test Chart(s)	Appropriate Input Data	Objective Characterisation	Assumed Subjective Relationship
Modulation Transfer Function (MTF)	<ul style="list-style-type: none"> • Sine-wave or Siemens Star charts • Straight or slanted edge signals • White noise signals • Dead-Leaves chart 	<ul style="list-style-type: none"> • Edge Spread Function (ESF), or • Input & output Fourier power spectra, or • Input & output signal modulation 	System signal transfer	Perceived sharpness and resolution, and perceived quality with respect to these attributes.
Noise Power Spectrum (NPS)	Uniform tone signal patches	Captured noise image	System noise	Perceived noisiness, and perceived quality with respect to noise.
Noise Equivalent Quanta (NEQ)	Refer to MTF and NPS	<ul style="list-style-type: none"> • MTF • NPS • Mean linear signal 	System signal-to-noise performance	Perceived quality with respect to sharpness, resolution and noise attributes.

Table 2.2 Summary of the most relevant performance measures concerning spatial image quality.

The CSF is a threshold function for human spatial visual sensitivity. Nevertheless, IQMs cascade it with the MTF and NPS as if it were a transfer function for all visuo-cognitive processes of spatial image quality judgement. Contextual visual models [8], [35] are expected to be more suitable for image quality modelling. Unlike traditional CSFs, they account for relevant scene-dependent effects of visual masking. This thesis investigates their use in IQMs.

Chapter 3 Image Quality Metrics

The plethora of image quality metrics (IQM), and their contrasting characteristics, reflects the recent expansion of image quality analysis applications. This thesis is concerned with two-dimensional (2D) spatial IQMs for still images. Relevant genres are reviewed critically from a capture systems development perspective. The *Signal Transfer Visual IQM (STV-IQM)* and *Multivariate Formalism IQM (MF-IQM)* genres are concluded most appropriate.

A novel framework is proposed for measuring spatial image quality while accounting for imaging system and human visual scene-dependency. This involves revising STV-IQMs and MF-IQMs to address the limitations of their current input parameters (discussed in Chapters 1 and 2). Frameworks for two further, novel IQMs are presented, based upon a similarly revised Noise Equivalent Quanta (NEQ) measure.

3.1 Image Quality Metrics (IQM) Review

Wang [202] states that objective image quality assessment represents a converging point of many research disciplines, including imaging systems development, signal and image processing, information theory, computer vision, visual psychophysics, machine learning and neural physiology. The diverseness of these approaches has evolved a broad spectrum of IQMs, developed for different applications.

This review evaluates a large number of IQMs that describe spatial image quality. They are classified by genres defined by the author that relate to their characteristics, input parameters, and intended applications. Each genre is analysed critically from the perspective of spatial image quality evaluation of 2D still image capture systems. The genres are summarised below and in Table 3.1. Most other IQM reviews evaluate metrics from a particular genre only, with respect to their intended application only [38], [52], [203]–[224].

The STV-IQMs, MF-IQMs and *Image Fidelity Metrics (IF-IQM)* are reviewed in Sections 3.1.1, 3.1.2 and 3.1.3, respectively. They employ system performance measures and human visual system (HVS) models from imaging and vision science, respectively. The most

relevant STV-IQMs to this thesis are multivariate (i.e. they account for multiple attributes, unlike univariate metrics that account for one attribute only). They are based upon fundamental relationships between imaging system signal-to-noise performance and perceived image quality; these relationships are discussed in Section 2.4. MF-IQMs apply the Minkowski combination to account for the perceived effect of several attributes/artefacts. IF-IQMs describe image fidelity specifically, either as the probability of the detection of differences between two images, or the perceived magnitude of these differences. Metrics from the more recent Computational IQM (CP-IQM) genre are reviewed in Section 3.1.4. They are generally black-boxes that analyse image content, features, or natural scene statistics (NSS); most apply top-down visual modelling. Further terminology that is used to classify IQMs in this thesis is defined below.

IQM Genre	Reference Type	Input Parameter(s)	Output Data	Correlating Data
Signal Transfer Visual IQM (STV-IQM)	No Reference	System performance measurements	Quality or sharpness score	Perceived image quality or sharpness ratings (JND)
Multivariate Formalism IQM (MF-IQM)	No Reference	System performance measurements	Quality score	Perceived image quality rating (JND)
Image Fidelity Metric (IF-IQM)	Full Reference	Test image + Reference image	Fidelity score or error map	Perceived image fidelity ratings (JND), or probabilities of detection of image differences
Computational IQM (CP-IQM)	Full / Reduced Reference	Test image + Reference image	Quality difference score	Differential Mean Opinion Score (DMOS)
	No Reference	Test image	Quality score	Mean Opinion Score (MOS)

Table 3.1 Summary of image quality metric (IQM) genres; JND refers to just-noticeable difference.

Full-Reference and Reduced-Reference IQMs compare information from a test and reference image (Figure 3.1). Full-Reference IQMs consider all information from these images; Reduced-Reference IQMs examine certain features only. Both aim to correlate with Differential Mean Opinion Scores (DMOS), or quality difference ratings in terms of just-noticeable difference (JND), where observers responded to the question “which image is preferable and to what extent?”

IF-IQMs are Full-Reference by default. *Threshold IF-IQMs* aim to correlate with the probability of detection of image differences, where observers responded to the question “can you see any differences between the images?” *Suprathreshold IF-IQMs* aim to correlate

with perceived image fidelity ratings in JND units, where observers responded to the question “what is the perceptual level of magnitude of differences between the images?”

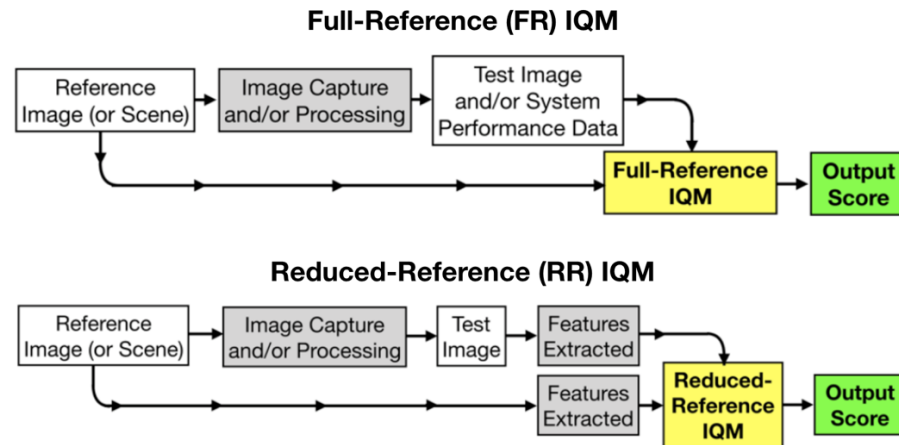


Figure 3.1 Generalisation of Full-Reference and Reduced-Reference image quality metric (IQM) characteristics.

No-Reference IQMs do not require data from a reference image (Figure 3.2). They aim to correlate with Mean Opinion Scores (MOS), or quality ratings in terms of JND, where observers responded to the question “how do you rate the level of goodness of the image?”

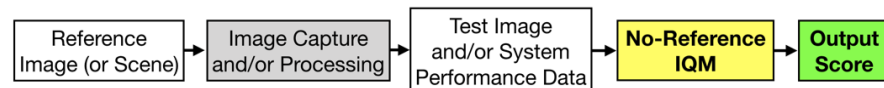


Figure 3.2 Generalisation of No-Reference (NR) image quality metric (IQM) characteristics.

The mentioned psychophysical JND ratings in terms of image quality, fidelity and sharpness relate to the probability of seeing a difference between two images [1, p. 355]; Section 2.5 defines 1 JND with respect to variation of the contrast of a stimulus. Psychometric scales relating to perceived image quality/fidelity/sharpness can be derived by various psychophysical paradigms [1, p. 356]: the paired comparison method, method of limits, method of adjustments, or the method of constant stimuli. The test image is a generally a distorted (or more distorted) version of the reference image. These psychophysical measurements are most relevant when they are output on ratio or interval scales. Ratio scales are the most advanced and have an origin equal to zero. Interval scales are floating scales. Equal differences in ratio/interval scale values are representative of equal perceptual differences in image quality or attribute strength.

MOS and DMOS refer to ratings of image quality and quality difference, respectively, that are not expressed in JNDs. They are assigned to images from the image quality assessment databases [225]–[241] used to validate and benchmark CP-IQMs. These images can contain varying levels of white, Gaussian, masked, quantisation, or spatially varying noise, or artefacts associated with non-content-aware denoising, Gaussian blur, Joint Photographic Experts Group (JPEG) and JPEG 2000 compression and transmission errors. The most popular databases were derived by 5-point psychophysical categorical scaling tests, converted linearly to the range 1 to 100 [242].

Varying levels of calibration are applied by default by each IQM to optimise correlation with their corresponding image quality rating dataset (as illustrated in Figure 3.3). Certain univariate STV-IQMs such as the Acutance (Equation 3.1) apply minimum or no use of calibration constants. This is because they describe fundamental relationships between image quality, system performance measures and/or bottom-up HVS models. IF-IQMs and multivariate STV-IQMs require low levels of calibration typically, for the following reasons. Multivariate STV-IQMs are based upon simple, well-established signal-to-noise relationships between physical parameters and image quality. IF-IQMs apply extensive bottom-up visual modelling and fidelity perception is more straightforward to model than quality perception. MF-IQMs require greater levels of calibration and some curve fitting since they are more complex and account for several attributes and artefacts. CP-IQMs involve the highest levels of calibration. Most examples apply curve fitting extensively.

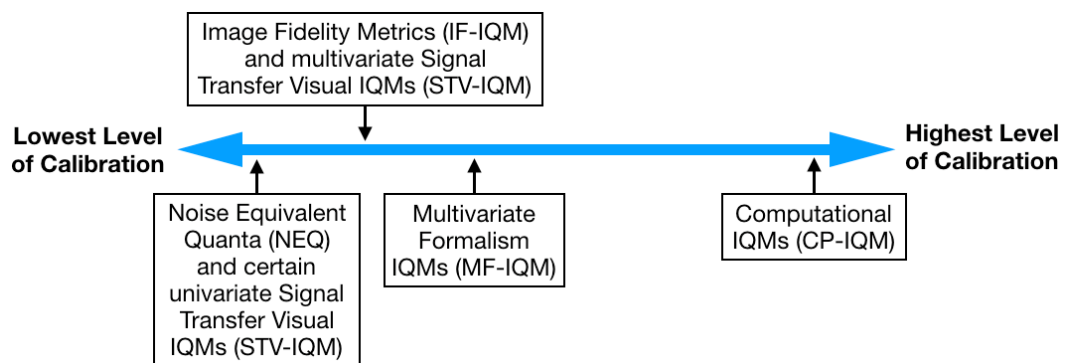


Figure 3.3 Approximate relative levels of calibration employed by image quality metrics (IQM) from different genres and the Noise Equivalent Quanta (NEQ) system performance measure.

3.1.1 Signal Transfer Visual IQMs (STV-IQM)

STV-IQMs have long been employed in the development and quality modelling of analogue and digital capturing systems. They have been reviewed comprehensively [38], [52], [213], [243]. Most univariate STV-IQMs [22, p. 54], [104], [196], [198], [199], [244]–[247] account for resolution and sharpness attributes. They cascade the Modulation Transfer Function (MTF) of system components (Equation 2.28) to model the imaging chain signal transfer. The contrast sensitivity function (CSF) is also cascaded to account for spatial HVS sensitivity. This cascade is integrated to yield a single figure sharpness (or quality) score.

The Acutance measure of the IEEE P1858 Camera Phone Image Quality (CPIQ) Standard [22, p. 54] (Equation 3.1) is a recent univariate STV-IQM. $CSF(u)$ is contrast sensitivity with respect to retinal spatial frequency in cycles/degree, u . The cut-off frequency is u_{max} . $MTF_{system}(u)$ and $MTF_{display}(u)$ are MTFs of the device and the display, respectively.

$$Q_T = \frac{\int_0^{u_{max}} MTF_{system}(u) \cdot MTF_{display}(u) \cdot CSF(u) du}{\int_0^{\infty} CSF(u) du} \quad (3.1)$$

The display MTF is modelled by a sinc function that describes the frequency response of an ideal pixel [22, p. 16]. The constant k_{disp} is 0.0243 for a high-quality display of 100 *pixels per inch (ppi)* resolution.

$$MTF_{display}(u) = \left| \frac{\sin(\pi k_{disp} u)}{\pi k_{disp} u} \right| \quad (3.2)$$

Multivariate STV-IQMs express image quality as the integral of a visually-weighted signal-to-noise calculation, building upon the founding work of Shannon [171], Schade [43], Bartleson [111], Nelson [248] and Higgins [136].

Barten's [12] *Square Root Integral with Noise (SQRIn)* [8], [33] is the most widely applied multivariate STV-IQM. It was validated successfully with respect to analogue capture system images [249] varying in contrast, luminance, resolution, addressability, viewing distance and field size [52], [250]. It has since been reformulated by Töpfer and Jacobson [169]. Under certain conditions, however, its JND scaling and correlation with perceived graininess were found to be of limited accuracy [169]. It was found to model the perceived quality of non-linear JPEG6b compressed images with limited success [9], [10], [251].

Töpfer and Jacobson's [169] Perceived Information Capacity (PIC) expands upon Shannon's Channel Capacity [171] and the related Information Capacity measure [173, p. 631] (defined by Equations 2.40 and 2.41, respectively) and accounts for bottom-up HVS modelling. Its calculation is comparable to the SQRI_n, with improved correlation with perceived graininess under low-contrast conditions and improved JND scaling [169].

Equations 3.3 and 3.4 define the PIC [169] and Töpfer and Jacobson's [169] reformulation of the SQRI_n [12], respectively, for analog capture systems. In this thesis, all further references to the SQRI_n refer to this reformulation unless otherwise stated. $S(u)$ and $N(u)$ are the signal power spectrum and Noise Power Spectrum (NPS) of the displayed print, respectively (given by Equations 3.5 and 3.6). $NPS_{visual}(u)$ is the neural NPS [12]. $CSF(u)$ is Barten's CSF [186] vs retinal spatial frequency, u , in cycles/degree. The maximum perceivable frequency is u_{max} . k_1 and k_2 are calibration constants.

$$PIC = k_1 \sqrt{\int_0^\infty \ln \left(1 + \frac{S(u)CSF^2(u)}{N(u)CSF^2(u) + NPS_{visual}(u)} \right) \frac{du}{u}} + k_2 \quad (3.3)$$

$$SQRI_n = \frac{k_1}{\ln 2} \int_0^{u_{max}} \left[\frac{S(u)CSF^2(u)}{N(u)CSF^2(u) + NPS_{visual}(u)} \right]^{0.25} \frac{du}{u} + k_2 \quad (3.4)$$

The signal spectrum of the displayed print from an analog system, $S(u)$, is expressed below [169]. $MTF_{film}(u)$, $MTF_{lens}(u)$ and $MTF_{paper}(u)$ are the MTFs of the photographic negative, enlarger lens, and paper, respectively. m is magnification and u is retinal spatial frequency in cycles/degree. $S_o(u)$ is the scene power spectrum modelled by Barten [12].

$$S(u) = S_o(u) \cdot MTF_{film}^2(m \cdot u) \cdot MTF_{lens}^2(m \cdot u) \cdot MTF_{paper}^2(u) \quad (3.5)$$

The corresponding NPS of the displayed print, $N(u)$, is given below [169]. $NPS_{paper}(u)$ and $NPS_{film}(u)$ are the NPSs of the photographic paper and film, respectively. $MTF_{paper}(u)$ and $MTF_{lens}(u)$ are the MTFs of the photographic paper and enlarger lens, respectively. γ_{paper} is the gamma of the paper, m is magnification, and u is retinal spatial frequency in cycles/degree. a_{nn} converts the density of the film to the printing density.

$$N(u) = NPS_{film}(m \cdot u) \cdot \gamma_{paper}^2 \cdot m^2 \cdot a_{nn}^2 \cdot MTF_{lens}^2(m \cdot u) \cdot MTF_{paper}^2(u) + NPS_{paper}(u) \quad (3.6)$$

The Effective Pictorial Information Capacity (EPIC) metric was developed by Jenkin *et al.* [195], [252]. It is based upon Shannon's [171] information capacity, like the Perceived Information Capacity [169]. It is computed as c in Equation 2.41 [136] as follows. The effective pixel size, n , is given by the width of the Line Spread Function (LSF) at approximately 50% of its peak value. This LSF is calculated [195] as the modulus of the inverse discrete Fourier transform (DFT) of the cascade of the imaging chain's MTF and the optical MTF of the eye (defined in Equation 2.57). The number of effective distinguishable signal levels, m , is modelled by the method of Altman and Zweig [253] according to the standard deviation of the noise of the system. Initial validations were carried out with respect to JPEG [195], [252] and JPEG 2000 [252] compressed images of scenes. Further validations with respect to images of scenes that varied in sharpness and noisiness found the EPIC correlated well with mean image quality ratings over several scenes, but not with ratings for individual scenes containing unusual structural content [254].

Multivariate STV-IQMs, in particular, are well suited for capture system design and quality modelling. They describe relationships between spatial image quality and real fundamental physical quantities related to imaging system performance, the HVSs' physiology and the viewing conditions. Their output scores are delivered by logical step-by-step calculations [185], [206] that establish causal justification for their relationship with perceived image quality. Their use of bottom-up spatial HVS models (defined in Section 2.5) makes them more accurate and appropriate than assumptions based only on signal transfer, noise, or signal-to-noise measurements (defined in Sections 2.3.2, 2.3.3 and 2.4, respectively).

STV-IQMs are more straightforward to implement and less computationally intensive than MF-IQMs and IF-IQMs. They are modular and easily revised [185]. Their input parameters can be altered independently [206] enabling convenient simulation of imaging chain and viewing condition variations. This is beneficial when optimising the trade-offs between cost and quality. STV-IQMs do not require information from the output image. Thus, future systems can be analysed speculatively without the need for functional prototypes, providing the MTF and NPS of their components are known. The SQRIn and PIC are of primary interest to this project since they implement the MTF and NPS directly.

Inaccuracies in MTF and NPS measurements from systems using non-linear content-aware processing are carried into the output scores of STV-IQMs (these inaccuracies are discussed in Sections 2.3.2 and 2.3.3, respectively).

The CSFs employed by STV-IQMs lack models for visual masking that are implemented in many IF-IQMs [34], [185], [255]. The CSF is also treated as an HVS transfer function, but it does not describe HVS processing of suprathreshold signals of all magnitudes [194] (assumptions regarding HVS linearity only apply close to threshold limits [183, p. 154], [256]). High-level cognitive processes associated with observer quality consciousness are also unaccounted for. They are discussed in Section 2.5.

The above are expected to cause the poor overall correlations between STV-IQMs and the subjective image quality of non-linear systems/processes [9], [10], [251] (e.g. Figure 1.5).

3.1.2 Multivariate Formalism IQMs (MF-IQM)

Keelan's [46] MF-IQM implements the Minkowski combination [257] that has been used traditionally as a visual perceptive model [258], [259], including in metrics for the combined perception of sharpness and graininess [111] and coding artefacts [260], [261]. Overall perceived image quality loss, QL_m , is modelled by combining calibrated univariate metrics for quality loss, QL_i , due to several attributes/artefacts, i (Equation 3.7). n_{max} is the power parameter (Equation 3.8) [46, p. 161]. Attributes (or artefacts) that are predicted to be of highest perceived magnitude dominate the overall quality loss score; this is in agreement with human perception [110], [111].

$$QL_m = \left(\sum_i (QL_i)^{n_{max}} \right)^{\left(\frac{1}{n_{max}} \right)} \quad (3.7)$$

The power parameter, n_{max} , adjusts the level of dominance of the most significant attribute [46, p. 162]. If $n_{max} = 2$, Equation 3.7 becomes a root mean square (RMS) sum representing the Euclidean distance. QL_{max} is the maximum quality loss for the system under the viewing conditions. The constants c_1 and c_2 are 2 and 16.9, respectively [3].

$$n_{max} = 1 + c_1 \cdot \tanh \left(\frac{QL_{max}}{c_2} \right) \quad (3.8)$$

The IEEE P1858 CFIQ Standard [22] defines several attribute metrics that are calibrated for the purpose. Each attribute metric expresses quality loss in JNDs with respect to one of the following: texture loss/blur, visual noise, edge Spatial Frequency Response (SFR), local geometric distortion, lateral chromatic displacement, chroma level and colour uniformity.

Figure 3.4 summarises the function of the MF-IQM for the two attribute metrics of greatest interest to this project, texture loss and visual noise. This MF-IQM is referred to from here on in this thesis as *the CPIQ metric*.

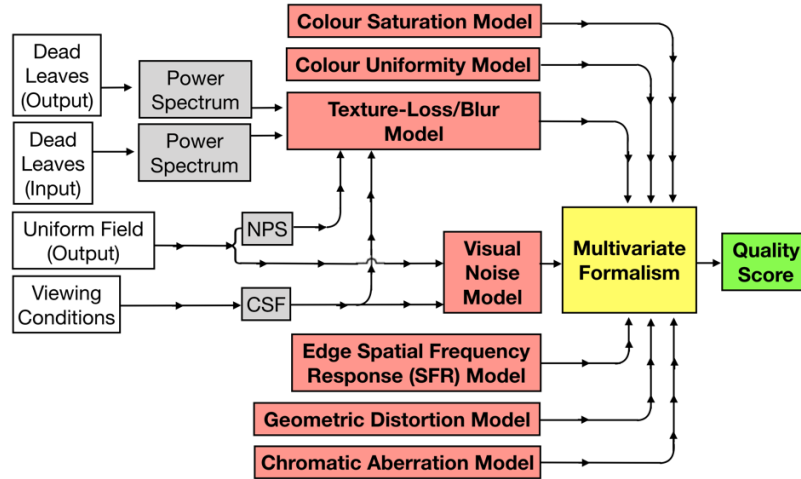


Figure 3.4 Diagram describing processing by the IEEE P1858 Camera Phone Image Quality (CPIQ) Standard's [22] implementation of Keelan's multivariate formalism IQM (MF-IQM) [46]. Input parameters are shown for the texture blur and visual noise attribute metrics only.

Subjective quality loss, $QL_{Texture_Loss}$, predicted by the *texture loss/blur attribute metric* is defined below [22, p. 56]. Q_T is the system Acutance (Equation 3.1) computed using the direct dead leaves MTF (Equation 2.31).

$$QL_{Texture_Loss} = 3.83 Q_F + 52.9 Q_F^2 + 9.97 Q_F^3, \quad (3.9)$$

where

$$Q_F = 0 \quad \text{for } Q_T > 0.961$$

$$Q_F = 0.961 - Q_T. \quad \text{for } Q_T \leq 0.961$$

Subjective quality loss, QL_{Visual_Noise} , according to the *visual noise attribute metric* [22, p. 46] is given below. Ω is a visual noise objective metric for the system (defined in Equation 3.11) [22, p. 69]. The constants a , b , and c are 0.319, 0.138 and 0.0049, respectively.

$$QL_{Visual_Noise} = \frac{\Omega - a}{b} - \frac{c \times \ln\left(1 + b \times \frac{\Omega - a}{c}\right)}{b^2} \quad \text{for } \Omega > 0.319 \quad (3.10)$$

$$QL_{Visual_Noise} = 0 \quad \text{for } \Omega \leq 0.319$$

The *visual noise objective metric* [22, p. 69] revises the metric of the same name from ISO 15739 [168]. It is computed using noise images from captured uniform-tone patches. These images are first linearised by an inverse standardised RGB (sRGB) tone transfer function [132]. Three colour space conversions are then applied: linear sRGB to CIE XYZ (D65), CIE XYZ (D65) to CIE XYZ (Illuminant E), and CIE XYZ (Illuminant E) to AC₁C₂. The resultant image is passed through a 2D Fast Fourier Transform (FFT) and filtered by the following functions: Johnson and Fairchild’s [188] luminance and chrominance CSFs (defined in Equations 2.54 and 2.55, respectively), the modelled display MTF [22, p. 16] (defined by Equation 3.2), and a frequency domain high pass filter. The filtered image is transformed back to the spatial domain by applying the inverse 2D FFT. Three further colour space conversions are applied: AC₁C₂ to CIE XYZ (Illuminant E), CIE XYZ (Illuminant E) to CIE XYZ (D65), and CIE XYZ (D65) to CIELAB [262]. Finally, the *total visual noise*, Ω , is given below [22, p. 44]. σ_L^2 , σ_a^2 and σ_b^2 are the variances of the L^* , a^* , and b^* components of the processed noise image, respectively, and σ_{La}^2 is its L^*a^* covariance.

$$\Omega = \log_{10}(1 + 23 \sigma_L^2 + 4.24 \sigma_a^2 - 5.47 \sigma_b^2 + 4.77 \sigma_{La}^2) \quad (3.11)$$

The recently standardised [22] and validated [3] CPIQ metric and its underlying MF-IQM [46] have proven to be accurate and versatile models for capture system image quality [178], [263]. The Minkowski combination at their core and their various attribute metrics are highly modular. Predictions of quality loss by each attribute metric, and the overall quality loss, are justified by logical step-by-step calculations using standard imaging system performance measures and bottom-up visual models. The Minkowski combination requires attribute metrics to be independent [46, Sec. 11.4], [194]. This is not always the case for the image quality attributes [162], [163], but this has not proved a problem in validation studies [3]. The CPIQ metric is more computationally sophisticated and demanding than STV-IQMs or CP-IQMs. For example, its visual noise attribute metric involves six colour space transformations and three stages of Fourier filtration. The CPIQ metric also requires more extensive calibration than STV-IQMs or IF-IQMs (including some curve fitting) to ensure its attribute metrics are perceptually “balanced”. Thus, its relation to the underlying physics of the imaging and visual system is less direct than these IQMs. The input parameters of each attribute metric can be altered independently. However, care should be taken not to disturb this calibration. The CPIQ metric’s texture loss and visual noise attribute metrics were found to be primary drivers of overall perceived image quality [3]. But both these

attributes can be accounted for, in theory, by the simpler, multivariate STV-IQMs. The visual noise attribute metric uses noise images obtained from imaged uniform patches that have been shown to describe system noise inaccurately in Figure 2.3. The employed Johnson and Fairchild luminance CSF [188] is conveniently uncomplicated but does not account for visual masking, the viewing conditions, the luminance of the stimulus, or the field size.

3.1.3 Image Fidelity IQMs (IF-IQM)

IF-IQMs model the perceptibility (fidelity) of image differences (defined in Section 2.1). They are often erroneously described as expressing perceived image quality difference (i.e. the acceptability of any perceived differences). Threshold IF-IQMs model the probability of observers discriminating differences between an ideal (reference) and distorted (test) image [264]. Suprathreshold IF-IQMs model the perceived magnitude of suprathreshold differences between these images [265]–[267]. Both sub-genres are Full-Reference and have been reviewed extensively [218]–[223].

Threshold IF-IQMs have mainly been employed to assess and optimise compression algorithms [255], [264], [268]–[274], displays and imaging systems [185], [264]. The test and reference images pass through several processing stages, generalised by Figure 3.5. These stages can account for display performance, amplitude non-linearity, divisive normalisation, opponent-space colour vision, contrast sensitivity and visual masking. Probability modelling is then applied to the differences computed between the images at each pixel. This yields distortion maps that express the probability of discrimination of these differences, which can be pooled to single figures [264].

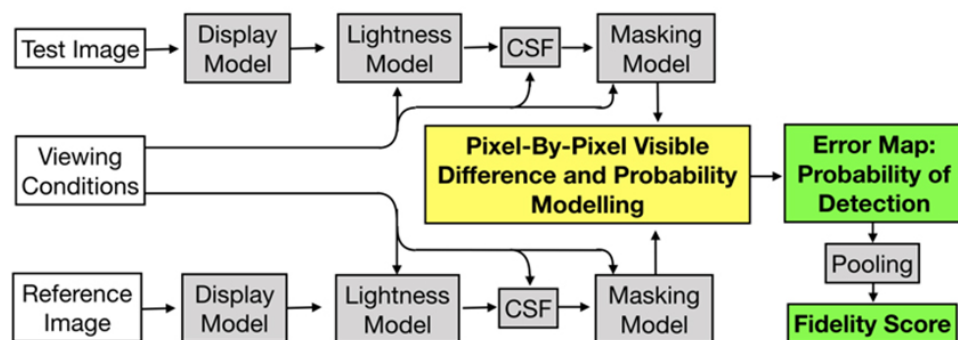


Figure 3.5 Generalisation of processing by threshold image fidelity metrics (IF-IQM), adapted from [264].

Suprathreshold fidelity metrics [36], [265]–[267], [275], [276] account for spatial vision by filtering opponent colour space images with threshold CSFs [188] or suprathreshold contrast discrimination models [277] (Figure 3.6). The popular Spatial Extension to CIELAB (S-CIELAB) [265]–[267] and comparable CSF/CIEDE2000 [36], [275] metrics then apply perceptual colour difference models from CIELAB [262] (ΔE_{ab}^*) and CIEDE2000 [278] (ΔE_{00}^*), respectively. S-CIELAB has since been revised [276] to account for chromatic adaptation and local contrast detection [279].

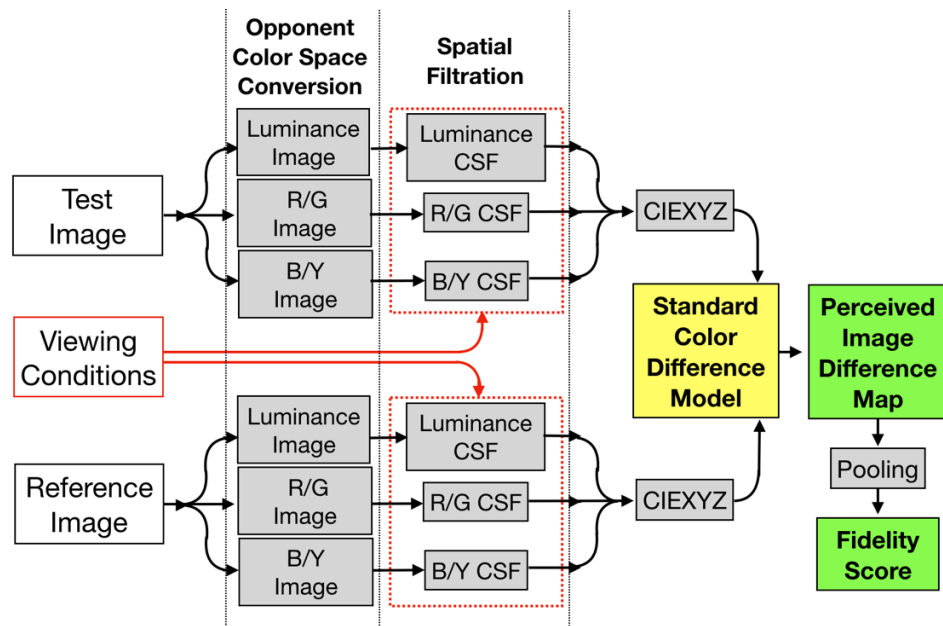


Figure 3.6 Generalisation of processing by suprathreshold image fidelity metrics (IF-IQM).

The modular, cascade-based IF-IQMs relate to the underlying physics of imaging systems and human vision and provide logical step-by-step pathways to the fidelity score. They generally use more comprehensive bottom-up HVS modelling than STV-IQMs and MF-IQMs. These models include mechanistic 2D CSFs for luminance and chrominance detection or discrimination that can account for visual masking [264] or selectivity by orientation [34], [52]. IF-IQMs are unsuitable for capture system image quality modelling, however, for the following reasons. They do not account for capture system performance measures. Reference images do not exist in image capture scenarios. Perceived quality and fidelity do not always correlate [36] (as discussed in Section 2.1). They are more computationally intensive and complex than both STV-IQMs and CP-IQMs since they perform pixel-by-pixel comparisons between two images after several stages of processing. Suprathreshold IF-IQMs apply threshold CSFs as HVS transfer functions, which has

limitations as stated in Section 3.1.1 [194]. Threshold IF-IQMs apply the CSF correctly as a threshold visual model but often clip the probability of detection of suprathreshold image differences to value 1 [34], [185]. Thus, they are best suited for optimising perceptually lossless algorithms [36] or assessing barely perceptible image differences [276]. There is no universally accepted method for pooling their 2D distortion maps into single figure measurements.

3.1.4 Computational IQMs (CP-IQM)

CP-IQMs have increased rapidly in number since 2004 along with the amount of image quality assessment applications. The following are notable examples of the many reviews available [203]–[212], [214]–[217], [224]. CP-IQMs can be characterised broadly as black-box models that establish statistical relationships between subjective MOS/DMOS datasets and objective image content, features, or NSS. Most examples apply top-down, hypothetical visual models. These models often aim to account for more complex, scene-dependent visual processes associated with image quality judgement.

Full-Reference and Reduced-Reference CP-IQMs aim to correlate with DMOS ratings. They include structural, mathematical, information theoretical, machine learning, visual attention, and hybrid approaches. No-Reference CP-IQMs are categorised as machine learning and information theoretical approaches and aim to correlate with MOS ratings. Each approach is summarised below.

The structural approach has been the most generally accepted Full-Reference paradigm [280] since the Structural Similarity (SSIM) Index and corresponding Mean Structural Similarity (MSSIM) [114] metric were introduced in 2004; both are based on the less stable Universal Quality Index (UQI) [281]. They assume structural distortions in the test image are the primary driver of perceived spatial image quality loss [114]. Structural distortions are defined, hypothetically [114], by reduced levels of local correlation with the reference image [282] after local variance (contrast) and luminance differences have been accounted for. Each stage of SSIM processing is summarised by Figure 3.7 and defined in Appendix B.

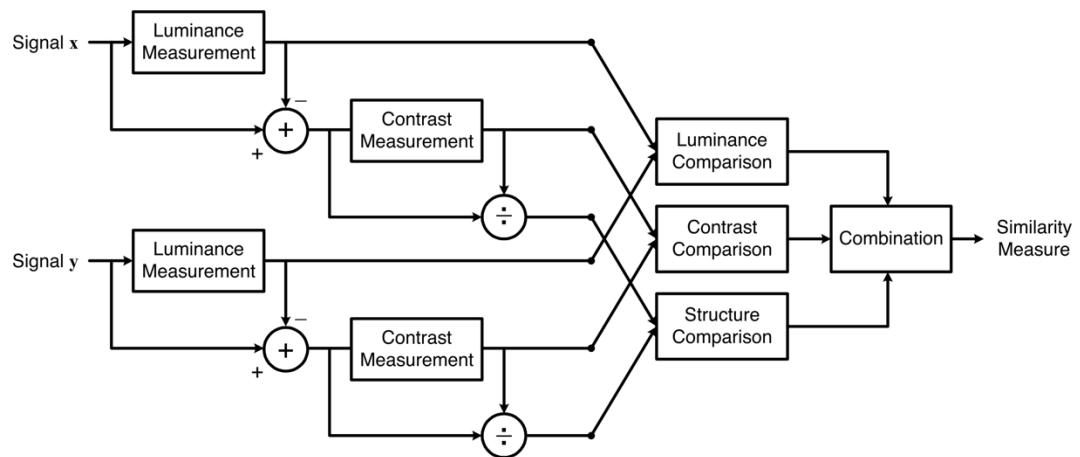


Figure 3.7 Diagram of processing by the Structural Similarity (SSIM) Index [114].

The SSIM has been extended to implement multi-scale [283], [284], wavelet domain [285] or Single Value Decomposition (SVD) [286], [287] processing or further classification of image structure [288], [289]. The limitations of structural approaches are as follows, Sheikh [126] states that there is no generally recognised way of defining perceived image structure or structural distortions. Haun and Peli [33] also describe the top-down HVS models employed as abstract and reduced, although some examples emulate visual masking.

The simplicity and tractability [290] of earlier mathematical approaches such as the peak signal-to-noise ratio (PSNR) and Mean Square Error (MSE) has ensured they are the most commonly applied CP-IQMs to date [291]. They employ no HVS modelling, however, and thus lack perceptual relevance [292] and consistency [126]. For example, they are sensitive to intensity scaling, rotation, translation, and zooming beyond perceptual levels [280]. They also describe low-magnitude distortions spread over large areas as more detrimental to quality than higher-magnitude distortions that are more locally concentrated; this does not agree with known perceptual behaviour. Other CP-IQM approaches extend mathematical approaches directly [100], [281], [293]–[296], such as many visual attention approaches.

Visual attention approaches apply saliency maps from eye-tracking experiments to weight distortion maps from other CP-IQMs, such as the MSE [100], PSNR [101], SSIM [101], [297] and the Visual Information Fidelity (VIF) [101]. Improvements in the accuracy of these IQMs were recorded after applying these weightings [100], [101], [297], particularly when gaze duration was accounted for [100]. Other research found minimal improvements [298], noting inaccuracies when perceptually relevant artefacts were in non-salient locations such as the sky and background [101], [298]. Weighting with algorithmically-generated

saliency maps (e.g. Figure 2.6) is less accurate or computationally efficient than with eye-tracking data [101]. Adaptive strategies are recommended to account for these limitations [101].

Reduced-Reference approaches [280], [299], [300] are relatively uncommon and compare extracted features such as edge maps or transform coefficients [280]. They are generally more computationally efficient than Full-Reference approaches, at the expense of accuracy.

Hybrid approaches are uncommon. Most examples merge bottom-up low-level HVS models with hypothetical models for higher-level processing [194], [208], [228], [232], [301]. Alternatively, they adapt existing Full-Reference or Reduced-Reference CP-IQMs to account for low-level vision [296].

Information theoretical approaches hold foundations in information theory [171] like the PIC [169] and EPIC [195] STV-IQMs. Their visual modelling, however, is top-down, abstract, and reduced [33]. The CSF and luminance masking models are not implemented, although the orientational selectivity of the HVS and divisive normalisation are accounted for. The Full-Reference Information Fidelity Criterion (IFC) [302] employs a wavelet-based orientation channel decomposition. The popular VIF [126] is a renowned No-Reference metric that builds upon the IFC.

Machine learning approaches treat image quality modelling as a supervised regression problem with respect to data from the test (and reference) image(s) and MOS (or DMOS) ratings [210], [303], [304]. Firstly, image feature(s) are extracted to narrow down the input data. These are usually distortion specific such as blurring, noise, texture, ringing, or blocking [305], [306]. The characteristics and optimal field of application of the metric are determined by which feature(s) are selected. The IQM is then trained to optimise (minimise) the difference between its estimated and desired outputs. Convolutional Neural Networks (CNN) are often used to solve this regression task [307]–[316]. Certain CNN-based examples require no distortion-specific training [314]–[316]. They can operate in the spatial domain [316] or adapt to different domains using supervised [315] or unsupervised approaches [314].

The CP-IQMs are accurate, computationally efficient and suitable for real-time image quality assessment applications. They fill the growing industry demand for user-friendly IQMs with minimal input parameters; unlike the other IQM genres, they do not demand prior

knowledge of system performance measures or psychophysical variables [317]. The fact that they are computed from information from the system's output image means they can often account for system performance scene-dependency and human visual scene-dependency.

However, CP-IQMs are less suitable for image capture systems development [206] for the following reasons. They lack modularity and do not employ standard imaging system performance measures. Their top-down, non-mechanistic visual models are mostly hypothetical, with limited reference to psychophysical measurements, the visual neurophysiology, or the viewing conditions (see Section 2.5 for further information). This unconstrained approach to image quality modelling has accelerated the evolution of CP-IQMs but also encourages short cuts when modelling complex physical and visuo-cognitive processes. These short cuts are particularly evident in the No-Reference IQMs that use CNNs to optimise correlation with subjective image quality datasets, with less regard for causation. The CP-IQMs have not been validated with respect to relevant capture system artefacts caused by non-linear content-aware demosaicing, denoising, sharpening, tone-mapping and/or compression, after accounting for lens blur, photon and sensor noise and Bayer sampling. Thus, if a benchmarking study finds that CP-IQMs correlate higher with MOS/DMOS than STV-IQMs or MF-IQMs, one cannot infer they are superior metrics for capture systems.

Full-Reference and Reduced-Reference CP-IQMs cannot account for visually enhancing distortions (e.g. sharpening) since the reference image is assumed to be ideal; reference images are also unavailable in capture system image quality assessment scenarios. No-Reference CP-IQMs are more suitable but are generally least mechanistic, modular, or accurate, with least relation to the physics of imaging systems or the HVS.

3.2 Proposed Image Quality Metric (IQM) Frameworks

The review concluded that MF-IQMs (Section 3.1.2) and multivariate STV-IQMs (Section 3.1.1) are the metric genres most applicable to spatial capture system image quality modelling. The SQRIn [12] and PIC [169] were identified as STV-IQMs that were directly suited to this project. Likewise, the CPIQ metric [22] was a suitable MF-IQM.

Frameworks to revise these three IQMs are presented in this section. Frameworks for two further IQMs are also defined (the *log Noise Equivalent Quanta (log NEQ)* and *Visual log*

NEQ), based upon a new *scene-and-process-dependent NEQ (SPD-NEQ)* measure. Each framework uses input parameters that aim to account for system signal transfer and noise scene-dependency, and spatial HVS scene-dependency. The rationale for using these parameters is given below.

Section 2.3.3 demonstrated that standard uniform patch NPS measurements are of limited relevance to capture systems using non-linear content-aware image signal processing (ISP). Three novel scene-and-process-dependent NPS (SPD-NPS) measures are presented in Chapter 4. They are calculated from *scene-and-process-dependent noise images* shown in Figure 2.3, in order to account for these limitations. These noise images are computed from repeated captures of images of any input scene, or an appropriate test chart, referred to from here on as *replicates*. Consequently, they account for the scene-dependent characteristics of temporally varying noise, caused by the application of non-linear ISP.

Section 2.3.2 discusses the limitations of applying MTFs measured from edges, sine-waves, white noise and even dead leaves signals, to capture systems using non-linear content-aware ISP. Chapter 5 presents three novel scene-and-process-dependent MTF (SPD-MTF) measures. These aim to address the above limitations by accounting for the scene-dependent signal transfer characteristics of non-linear ISP algorithms. They are calculated by revising the direct dead leaves MTF implementation [19] (Equation 2.31) to use SPD-NPS measures and DFT power spectra computed either from images of scenes or from suitable test charts.

Section 2.5 concluded the Contextual CSF (cCSF) [35] and Contextual Visual Perception Function (cVPF) [8] should be more appropriate HVS models for STV-IQMs and MF-IQMs than the traditional CSFs used presently. This is because they account for visual masking from relevant suprathreshold signals and noise. This thesis investigates the use of the cCSF or cVPF in each IQM proposed. Note that each metric is defined in this chapter with respect to these visual functions but can also be computed using the Barten CSF [183, p. 36], [186].

3.2.1 Log Noise Equivalent Quanta (log NEQ), Visual log NEQ, and Scene-and-Process-Dependent NEQ (SPD-NEQ)

The SPD-NEQ revises the NEQ (defined in Equation 2.49) to account for system signal transfer and noise scene-dependency. The one-dimensional (1D) SPD-NEQ is defined below and described by Figure 3.8. $MTF_{SPD}(u)$, is the SPD-MTF, $NPS_{SPD}(u)$ the SPD-NPS, μ the mean linear signal, and u is spatial frequency in cycles/degree.

$$NEQ_{SPD}(u) = \frac{MTF_{SPD}^2(u)}{NPS_{SPD}(u)/\mu^2} \quad (3.12)$$

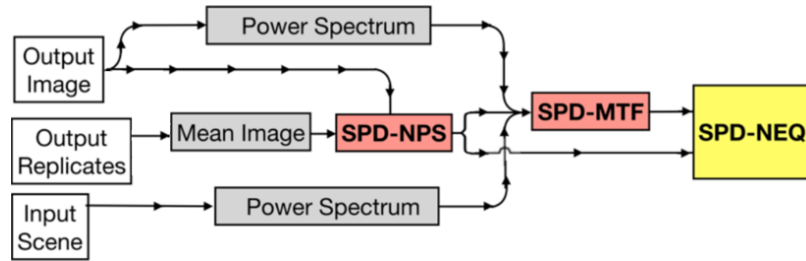


Figure 3.8 Diagram describing processing by the scene-and-process-dependent Noise Equivalent Quanta (SPD-NEQ) system performance measure.

The log NEQ metric is defined by Equation 3.13. It provides a single figure image quality score by taking the logarithm of the integral of an appropriately weighted SPD-NEQ measure. The Visual log NEQ metric (Equation 3.14) accounts for the spatial sensitivity of the HVS but is otherwise identical to the log NEQ. In both equations, $MTF_{Display}(u)$ is the MTF of the display [22, p. 16] (Equation 3.2), $NEQ_{SPD}(u)$ is the SPD-NEQ (Equation 3.12), u is the retinal spatial frequency in cycles/degree, u_{max} is the maximum perceivable frequency. $CSF_{Contextual}(u)$ is either the cCSF [35] (Equation 2.59) or the cVPF [8] (Equation 2.60). k_1 and k_2 are calibration constants concerning the gradient (gain) and intercept (offset) of the metrics.

$$LogNEQ = k_1 \log_{10} \left(\int_0^{u_{max}} MTF_{display}^2(u) NEQ_{SPD}(u) \frac{du}{u} \right) + k_2 \quad (3.13)$$

$$LogNEQ_{visual} = k_1 \log_{10} \left(\int_0^{u_{max}} CSF_{Contextual}^2(u) MTF_{display}^2(u) NEQ_{SPD}(u) \frac{du}{u} \right) + k_2 \quad (3.14)$$

Both the log NEQ and Visual log NEQ involve minimal levels of calibration. Therefore, they relate directly to the fundamental NEQ signal-to-noise relationship defined in Section 2.4. For both metrics, taking the logarithm of the integral was justified empirically. It caused them to predict observer image quality ratings from Chapter 6 more accurately than when the logarithm was not taken. Taking the logarithm also means they follow Fechner's law, which states that the relationship between the intensity of a stimulus and its perceived intensity is logarithmic. It should also be noted that the logarithm of a similarly weighted signal-to-noise function is taken by both the photographic information capacity (Equation

2.41) and the PIC (Equation 3.3), prior to integration. Although Shannon's channel capacity does not aim to describe visual perception specifically, it also takes the logarithm of the signal energy divided by the noise energy (Equation 2.40).

Simplicity and functionality were prioritised when developing the log NEQ and Visual log NEQ. Simple, functional metrics are used disproportionately over more complex metrics by the industry, as the Pareto principle would suggest [318]. Consequently, neither metric accounts for display or visual noise, since these affect perceived image quality less than display signal transfer (MTF) and spatial visual sensitivity (CSF), respectively, for current displays at most output luminance levels. In order for these parameters to have been accounted for by the log NEQ and Visual log NEQ, the input parameters to the SPD-NEQ would have needed to be specified separately in both metrics. This would have increased the metrics' complexity significantly (making them similar to the SQRIn and PIC).

3.2.2 Revised STV-IQMs

The revised SQRIn [12] is calculated using Equation 3.4 when the signal spectrum of the displayed image, $S(u)$, and the total imaging system noise, $N(u)$, are given by Equations 3.15 and 3.16, respectively. The CSF parameter, $CSF(u)$, denotes either the cCSF [35] (Equation 2.59) or cVPF [8] (Equation 2.60). $PS_{scene}(u)$ is the DFT power spectrum of the image of the scene (Equation 2.30). $MTF_{SPD}(u)$ and $NPS_{SPD}(u)$ are the SPD-MTF and SPD-NPS of the capture system, respectively. $MTF_{display}(u)$, $NPS_{display}(u)$, and $\gamma_{display}$ are the MTF [22, p. 16] (Equation 3.2), NPS (Equation 2.36), and gamma of the display, respectively. All other parameters of Equation 3.4 are as previously described.

$$S(u) = PS_{scene}(u) \cdot MTF_{SPD}^2(u) \cdot \gamma_{display}^2 \cdot MTF_{display}^2(u) \quad (3.15)$$

Similarly, the revised PIC [169] is computed by Equation 3.3, when $S(u)$ and $N(u)$ are given by Equations 3.15 and 3.16 as described above. $CSF(u)$ refers to either the cCSF [35] or cVPF [8]; all other parameters are as previously described. Figure 3.9 describes the input parameters and sub-processes for the revised SQRIn and PIC metrics.

$$N(u) = NPS_{SPD}(u) \cdot \gamma_{display}^2 \cdot MTF_{display}^2(u) + NPS_{display}(u) \quad (3.16)$$

$$\approx NPS_{SPD}(u) \cdot \gamma_{display}^2 \cdot MTF_{display}^2(u)$$

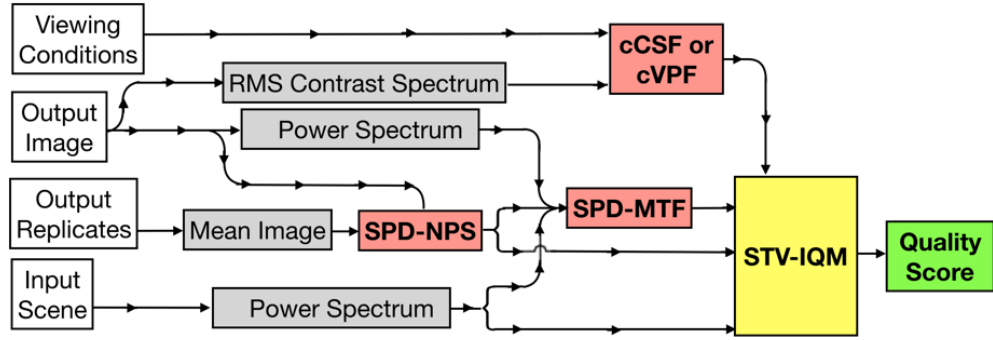


Figure 3.9 Diagram describing processing by both the revised SQRIn [12] and PIC [169] metrics.

3.2.3 Revised Camera Phone Image Quality (CPIQ) Standard Metric

The revised CPIQ metric is defined below in four stages.

A *scene-and-process-dependent texture acutance metric* is computed, Q_{SPD} (Equation 3.17). $MTF_{SPD}(u)$ is the SPD-MTF, $CSF_{Contextual}(u)$ is the cCSF [35] (Equation 2.59) or cVPF [8] (Equation 2.60), $MTF_{display}(u)$ is the MTF of the display [22, p. 16] (Equation 3.2), u is retinal spatial frequency in cycles/degree and u_{max} is the cut-off frequency. It is based on the texture acutance metric from the IEEE P1858 CPIQ standard (Equation 3.1) [22, p. 54].

$$Q_{SPD} = \frac{\int_0^{u_{max}} MTF_{SPD}(u) \cdot MTF_{display}(u) \cdot CSF_{Contextual}(u) du}{\int_0^{\infty} CSF_{Contextual}(u) du} \quad (3.17)$$

A *scene-and-process-dependent visual noise objective metric*, Ω_{SPD} , is computed by revising the equivalent metric from the IEEE P1858 CPIQ standard (Equation 3.10) [22, p. 46], as follows. The input noise image is a scene-and-process-dependent noise image (defined in Section 4.1). The luminance channel of each processed noise image is filtered with the cCSF [35] (Equation 2.59) or cVPF [8] (Equation 2.60), instead of Johnson and Fairchild's CSF [188] (Equation 2.54).

Q_{SPD} and Ω_{SPD} are substituted in place of Q_T and Ω in Equations 3.9 and 3.10, respectively, to yield revised attribute metrics for texture loss, $QL_{Texture_Loss}$, and visual noise, QL_{Visual_Noise} . All other parameters remain as defined previously.

Finally, a revised metric of overall quality loss (QL_m) is calculated with respect to the above revised attribute metrics using Equation 3.7; all other parameters are as previously described.

Figure 3.10 describes processing by the revised CPIQ metric. Note that the remaining five attribute metrics shown in Figure 3.4 can also be incorporated when calculating QL_m .

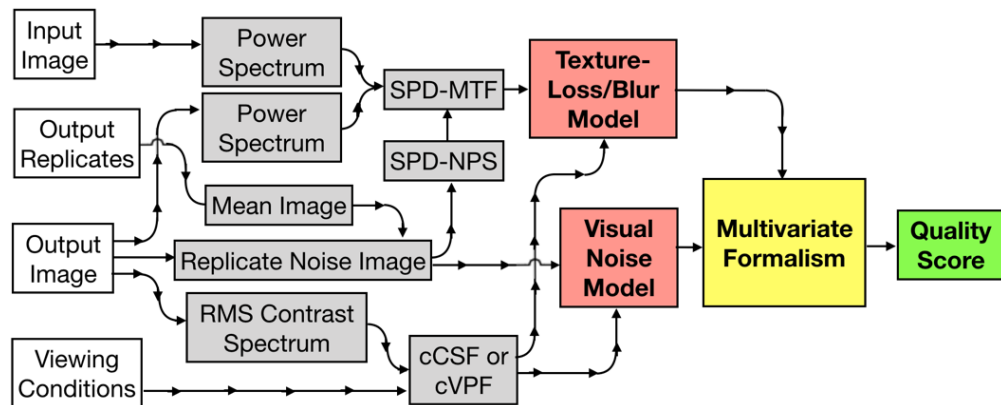


Figure 3.10 Diagram describing processing by the revised CPIQ metric [22] (with respect to texture blur and visual noise attribute metrics only).

3.3 Summary

Spatial IQMs have been classified by genre and reviewed critically from a capture systems development viewpoint. The MF-IQM and multivariate STV-IQM genres were concluded as most applicable to this project. The CPIQ metric [22], SQRIn [12] and PIC [169] were selected as suitable IQMs from these genres.

A novel image quality measurement framework was proposed that accounts for imaging system and human visual scene-dependency by using scene-dependent input parameters. Following this framework, the three IQMs selected above were revised to implement SPD-NPS and SPD-MTF system performance measures (defined in Chapter 4 and Chapter 5, respectively). They also applied either the cCSF [35] or cVPF [8] as visual models, although traditional CSFs can also be used. Two further, novel IQMs were proposed based upon a similarly revised NEQ framework. Chapter 6 validates each of these IQM frameworks.

Chapter 4 Validation of Scene-and-Process-Dependent NPSs (SPD-NPS)

The *scene-and-process-dependent Noise Power Spectrum (SPD-NPS) framework* is defined in this chapter. It characterises temporally varying capture system noise with respect to any input signal, accounting for system scene-dependency. Three novel SPD-NPS measures are defined that apply this framework. Each characterises one of the following:

- 1) capture system noise with respect to an image of a given input scene;
- 2) the average real-world noise of the system with respect to an image dataset comprising of different scenes depicting a variety of subjects;
- 3) system noise with respect to the dead leaves chart with an “average scene spectrum”.

Each measure is validated by evaluating measurements from image capture simulation pipelines, applying linear and non-linear image signal processing (ISP). Finally, a novel measure for capture system noise scene-dependency is presented and validated.

4.1 Derivation of the SPD-NPS Measures

The Noise Power Spectrum (NPS) is defined in Section 2.3.3. It is commonly measured from captured uniform tone patches. For systems applying non-linear ISP algorithms, however, Figure 2.3 demonstrates that noise in captured patches is often unrepresentative of temporally varying noise in captured scenes. For such systems, the latter forms structured patterns that are dependent on both the local scene content and the ISP algorithms used (i.e. the noise is local-content-dependent and scene-dependent).

Non-linear content-aware denoising influences this behaviour in particular. Uniform patches provide ideal conditions for these algorithms to remove noise (Section 2.2.1). Thus, for systems that apply non-linear denoising, noise images derived from uniform patches are

expected to underestimate noise in a given captured scene, as well as the average level of system noise in real-world capture scenarios. This renders uniform patch NPS measurements unsuitable for image quality modelling (as discussed in Section 2.3.3).

The SPD-NPS measuring framework is described by Figure 4.1. The framework uses a more appropriate scene-and-process-dependent noise image. Equation 2.39 computes this RGB noise image when $g(x, y)$ is a captured frame of any scene/target, and $\bar{g}(x, y)$ is the mean image of N registered replicates of the scene/target, $r_i(x, y)$, given by Equation 4.1. These noise images are computed using ten replicates in this thesis. This number was found to be adequate in related prior art [161]. Averaging the replicates cancels out temporally varying, random noise, meaning $\bar{g}(x, y)$ contains only the image signal and capture system fixed pattern noise (FPN). Thus, subtracting $\bar{g}(x, y)$ from $g(x, y)$ yields a noise image containing temporally varying noise only. The two-dimensional (2D) SPD-NPS is then computed using Equation 2.36, where $I(x, y)$ is the luminance component of the scene-and-process-dependent noise image; other parameters are as previously defined. The 1D SPD-NPS is given as the rotational average of this 2D SPD-NPS.

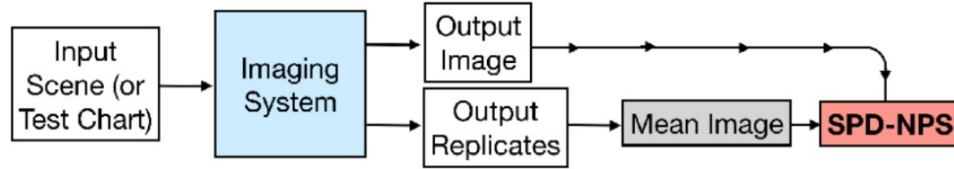


Figure 4.1 The scene-and-process-dependent Noise Power Spectrum (SPD-NPS) framework.

$$\bar{g}(x, y) = \frac{1}{N} \left(\sum_{i=1}^N r_i(x, y) \right) \quad (4.1)$$

Three SPD-NPS measures from this thesis implement the SPD-NPS framework. Each measure is defined below and summarised in Table 4.1 alongside the uniform patch NPS. They are further used in the scene-and-process-dependent Modulation Transfer Function (SPD-MTF) measures, presented in Chapter 5, and in the updated and novel image quality metrics (IQM) defined in Chapter 3. Note that from here on in this thesis, when the term noise is used in the context of an imaging system, it refers to the temporally varying noise of the system only, unless otherwise specified.


NPS measure	Input parameters	Summary of characterisation	Accounts for system noise scene-dependency	Sensitivity to system noise scene-dependency
Pictorial Image SPD-NPS	Single pictorial image N x pictorial image replicates	Characterises temporally varying noise with respect to a specific image	Yes	<div style="text-align: center;"> Highest  Lowest </div>
Mean Pictorial Image SPD-NPS	Multiple pictorial images N x replicates for each image	Characterises average real-world temporally varying system noise	Yes	
Dead Leaves SPD-NPS	Single dead leaves image N x dead leaves replicates	Estimates average real-world temporally varying system noise	Partially	
Uniform Patch NPS	Single uniform patch image	Characterises average real-world system noise	No	

Table 4.1 Summary of the uniform patch Noise Power Spectrum (NPS) and the scene-and-process-dependent NPS (SPD-NPS) measures. N is the number of replicate captures of each scene/chart.

The *pictorial image SPD-NPS* implements the SPD-NPS framework with respect to an image of any input pictorial scene. It can also be computed with respect to any static real scene. It accounts for the effect of the scene content on temporally varying noise, produced by systems applying non-linear content-aware ISP. Thus, it is the only measure capable of characterising noise in such systems with respect to a given input scene. It is expected to be a valuable input parameter for IQMs designed for non-linear systems. This is because it aims to compensate for the scene-dependent gain and offset in correlations between observer image quality ratings and current IQM output scores, as exemplified by Figure 1.5.

The *dead leaves SPD-NPS* implements the SPD-NPS framework with respect to the dead leaves chart that models the average power spectrum of natural scenes [17]. By measuring noise with respect to this mathematically generated “average scene signal”, the dead leaves SPD-NPS provides a convenient approximation of the average real-world level of temporally varying system noise. It is expected to be a more appropriate measure than the uniform patch NPS measure for systems applying non-linear content-aware ISP. However, assumptions associated with dead leaves Modulation Transfer Function (MTF) measurements also apply to it and may affect its accuracy (these assumptions are discussed in Section 2.3.2).

Two other noise measures derived from dead leaves signals have been described by Artmann [18]. The first determines the level of system noise indirectly, by subtracting the system’s intrinsic dead leaves MTF from its original dead leaves MTF [18]; strictly speaking, it is not an NPS measure. The second assumes that, the power spectrum of a hypothetical “noiseless captured test chart image” can be simulated by cascading the dead leaves test chart’s power spectrum with the intrinsic dead leaves MTF [18]. The NPS of the system is then measured

indirectly by subtracting the former power spectrum from the captured test chart's power spectrum [18]. Errors may result in both measures from the intrinsic MTF not accounting for the effects of reversible image processing [161] (as noted in Section 2.3.2), which is used by many current capture systems. The effects of such reversible processes are taken into account in the dead leaves SPD-NPS, and the other SPD-NPS measures presented in this thesis.

The *mean pictorial image SPD-NPS* is the only current measure for characterising the average real-world level of temporally varying system noise, accounting for the system's processing of real images of scenes, or the scenes themselves. It accounts for system scene-dependency. It is derived as the mean of a number of pictorial image SPD-NPS measurements, across a large and representative set of images of different scenes (i.e. it is given by $\bar{F}(s)$ in Figure 1.4(b), if $F(s_1)$ to $F(s_n)$ are SPD-NPSs for pictorial scene images s_1 to s_n , respectively).

It is unconventional to average over NPSs, or MTFs, in this fashion. However, Jenkin [2, p. 130] has demonstrated that averaging over MTFs mitigated inaccuracies caused by simulated quantisation noise provided that variations in the output modulation were distributed evenly around the input signal modulation. Equations 4.2 to 4.5 adapt the derivation of Jenkin's method [2, p. 130] to express how the mean pictorial image SPD-NPS tends toward the hypothetical "correct average real-world SPD-NPS" of the system as the scene image set increases in size, provided that comparable conditions are met.

Let the relationship between the pictorial image SPD-NPS, $N_{SPD}(u)$, the hypothetical "correct average real-world SPD-NPS" of the system, $N_{General}(u)$, and the variation between them, $\Delta N_{SPD}(u)$, be expressed below for one image of a scene. u is spatial frequency.

$$N_{SPD}(u) = N_{General}(u) \pm \Delta N_{SPD}(u) \quad (4.2)$$

When pictorial image SPD-NPS measurements for each image (i) in a set of p images of different scenes are averaged, the mean pictorial image SPD-NPS, $N_{Mean_SPD}(u)$, is expressed as:

$$N_{Mean_SPD}(u) = \frac{1}{p} \left[\sum_{i=1}^p N_{General_i}(u) \pm \Delta N_{SPD_i}(u) \right] \quad (4.3)$$

This summation is rewritten in expanded form:

$$N_{Mean_SPD}(u) = \frac{1}{p} \sum_{i=1}^p N_{General_i}(u) + \frac{1}{p} \sum_{i=1}^p \pm \Delta N_{SPD_i}(u) \quad (4.4)$$

When the limit of $p \rightarrow \infty$ is considered, the first and second term tend toward $N_{General}(u)$ and zero, respectively, provided that $\Delta N_{SPD_i}(u)$ is distributed evenly about zero. Thus, if the pictorial image SPD-NPS for a number of images of different scenes, $N_{SPD_i}(u)$, is distributed evenly around $N_{General}(u)$, the mean pictorial image SPD-NPS, $N_{Mean_SPD}(u)$, tends toward the average real-world SPD-NPS of the system, $N_{General}(u)$.

$$N_{Mean_SPD}(u) \lim_{p \rightarrow \infty} = \frac{1}{p} \sum_{i=1}^p N_{General_i}(u) + \frac{1}{p} \sum_{i=1}^p \pm \Delta N_{SPD_i}(u) \rightarrow N_{General}(u) \quad (4.5)$$

Provided the scene image set is large enough and reflects the characteristics of commonly captured scenes, the mean pictorial image SPD-NPS accounts for the effect of scene-dependent system behaviour on the average real-world system noise power. If the individual pictorial image SPD-NPS measures are biased, however, this bias is carried into the mean measure.

The SPD-NPS measures, and particularly the mean pictorial image SPD-NPS, are more computationally complex and more complicated to implement than the uniform patch NPS. Reducing the number of replicates lowers computational demand but causes underestimation of system noise. All replicates must be registered to mitigate bias that results from the scene- and process-dependent noise image being calculated inaccurately. Comparable practical measurements of noise from a smartphone [7] and digital single-lens reflex (DSLR) [161] camera required captured replicates to be registered using software, despite the use of a tripod [7], [161] and camera remote [7] during capture.

The SPD-NPS measures do not account for FPN which appears in all replicates, $r_i(x, y)$, and is thus present in the mean image, $\bar{g}(x, y)$, in Equation 4.1. For the same reasons, certain fixed patterns of artefacts that are caused by the mosaicing-demosaicing process are also

unaccounted for (shown in Figure 4.2). It should be noted that the uniform patch NPS also fails to account for such artefacts, since the artefacts only manifest when spatially varying input signals are captured. These artefacts are caused by the sparse sampling of information in fixed patterns by the Colour Filter Array (CFA), and the limited capabilities of demosaicing algorithms to recover the lost information. The intensity of these artefacts relates to the local gradient of each colour channel and is thus scene-dependent. Equation 4.6 isolates the artefacts, $N_{Fixed}(x, y)$, by modifying the image capture simulation pipelines described in Section 4.2. $I_{Blurred}(x, y)$ is the image of the scene after lens blur has been modelled. Bayer array sampling and demosaicing were then applied to this image, yielding an image corrupted with the mentioned artefacts, $I_{Corrupted}(x, y)$.

$$N_{Fixed}(x, y) = I_{Corrupted}(x, y) - I_{Blurred}(x, y) \quad (4.6)$$



Figure 4.2 Fixed patterns of artefacts caused by demosaicing the ‘Students’ image [7] (left), and the detail of it (right). Global contrast was enhanced to increase artefact intensity.

A final measure, the *pictorial image SPD-NPS standard deviation*, describes the level of scene-dependent variation in measurements of the system’s temporally varying noise. This variation is either caused by genuine system scene-dependency, or scene-dependent variation in the level of measurement error of the pictorial image SPD-NPS. It is the only current measure for system noise scene-dependency. It is computed by Equation 4.7 as the standard deviation, $s(u)$, of the pictorial image SPD-NPSs, $N_{SPD_i}(u)$, from a set of p images of different scenes, where u is spatial frequency. $N_{Mean_SPD}(u)$ is the mean pictorial image

SPD-NPS. Like the mean pictorial image SPD-NPS, this scene-dependency measure becomes more representative as p increases, provided its input parameters are unbiased.

The average real-world level of system noise and its level of scene-dependency are described by subtracting and adding the pictorial image SPD-NPS standard deviation from the mean pictorial image SPD-NPS (as demonstrated by the broken lines in Figure 4.6).

$$s(u) = \sqrt{\frac{1}{p-1} \sum_{i=1}^p (N_{SPD_i}(u) - N_{Mean_SPD}(u))^2} \quad (4.7)$$

4.2 Validation Methodology

The SPD-NPS measures were validated by analysing measurements from two simulated image capture pipelines, generated in MATLABTM (Figure 4.3). Controllable outputs were obtained from each pipeline after demosaicing, denoising and sharpening ISP was applied under a variety of simulated exposure conditions, with respect to 50 input images of scenes. This enabled each measure to be computed and compared after each ISP stage to evaluate pipeline scene-dependency and measurement accuracy. The pipelines were also used to validate the SPD-MTF measures and IQM frameworks in Chapters 5 and 6, respectively.

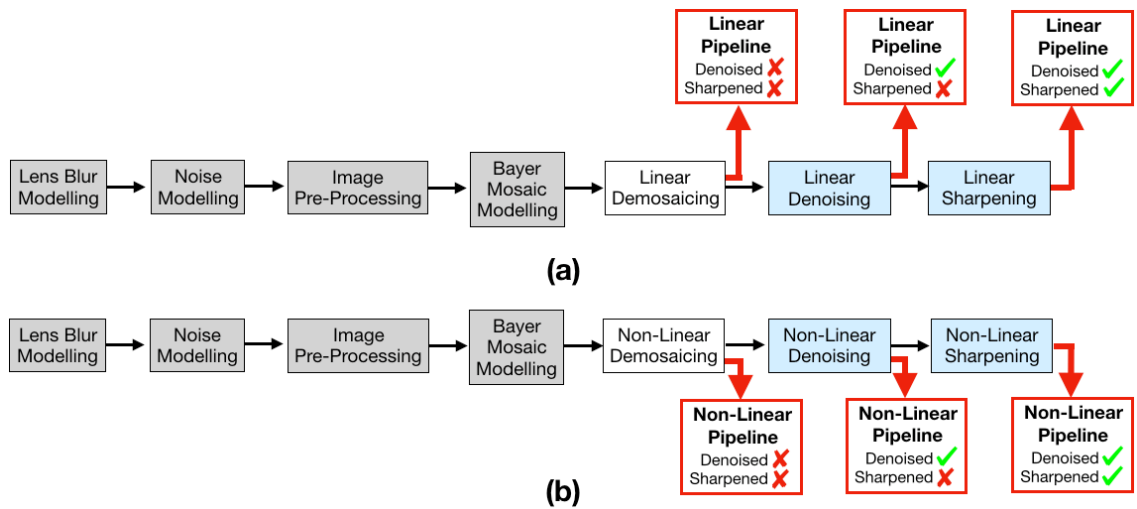


Figure 4.3 The linear (a) and non-linear (b) camera simulation pipeline modelling and image signal processing (ISP) stages. Output images were generated at points indicated by the red arrows.

Modelling that was common to both pipelines is shaded grey. ISP stages for which the linear and non-linear algorithms had very different effects on image quality to one another are shaded blue.

Both pipelines implemented identical modelling of physical camera processes, including lens blur, photon and sensor noise, image pre-processing and Bayer array sampling. One pipeline then applied linear demosaicing, denoising and sharpening ISP. The other pipeline applied the equivalent non-linear content-aware ISP. These pipelines are referred to as the *linear pipeline* and the *non-linear pipeline*, respectively, in this thesis.

The physical processes of digital image capture, described below, were modelled first in both pipelines. These processes are shown in grey in Figure 4.3. Their modelling did not aim to replicate a particular real system or simulate every relevant process on a quantum level. However, it produced images with comparable blur and noise artefacts to real physical systems for a range of exposure conditions. These images were then processed by published ISP algorithms. Note that it is the interactions between such ISP algorithms and the image signals and noise that are of most interest to this thesis, as well as the capabilities of the proposed measures/metrics to account for them.

Lens blur was modelled by convolving the input scene image with a Gaussian approximation for the central lobe of a diffraction-limited lens' airy disk [319], according to the f-number and pixel pitch of an iPhone 6 smartphone [320], [321]; outer airy rings were ignored. The peak wavelengths for the R, G and B channels were 570nm, 550nm, and 450nm, respectively. The Point Spread Function (PSF) kernel dimensions were 9-by-9 pixels.

2D Poisson noise was modelled using the `Imnoise` MATLABTM function [322] at maximum linear signal-to-noise ratios (SNR), \sqrt{q} , of 5, 10, 20 and 40, where q is the average number of events according to Equation 2.32. Thus, the SNR was computed at a normalised fractional linear luminance of 1 (i.e. at full-well capacity). These SNRs represent excellent (SNR 40) to very poor exposure conditions (SNR 5). The reduced quantum efficiency of blue and red pixels was modelled by scaling down the number of events in the R and B channels by factors of 2 and 3.33, respectively [165]. Read noise and dark current noise were modelled in one step, by adding limited levels of Gaussian noise to each channel with increased mean and variance at lower SNRs. This accounted for the fact that DSNU is dependent on exposure time (which was assumed longer at lower SNRs). FPN was not modelled.

The following pre-processing was then applied, in order:

- 1) The image was delinearised by applying a Gain Offset Gamma (GOG) model for the opto-electronic conversion function (OECF) with the gain, offset and gamma set to 1, 0 and 0.4545, respectively.
- 2) The gain was increased slightly.
- 3) Black and white level adjustments were applied using the `imadjust` MATLAB™ function, to remove the noise floor and recover highlights, respectively.

The gain, black and white level adjustments were more intensive at lower SNRs.

Pixel information was sampled from the R, G and B channels of the image according to a ‘grbg’ Bayer CFA. Most capture system simulations implement Bayer array sampling before noise modelling. The chosen order, however, yielded identical output images to this order and facilitated the independent scaling of noise in each colour channel.

The pipelines then implemented the ISP algorithms described below. If constants are quoted, the first constant is with respect to the pipelines tuned at a Poisson noise SNR of 5 and the second at SNR 40. Some capture systems can perform demosaicing, denoising and sharpening using a single algorithm, or two algorithms. However, the pipelines used three separate algorithms so that the effect of each process could be characterised in isolation. Algorithms capable of performing two of the above processes were restricted to perform one process only.

The linear pipeline used the following algorithms. Demosaicing was by the Malvar *et al.* [30] algorithm. It estimates the luminance gradient across each real CFA pixel after bilinear interpolation, with respect to the interpolated neighbouring pixels. The two interpolated colour channels at the same pixel location are adjusted according to this luminance gradient. This preserves image edges. The filter is linear since its region of support is within a 5x5 pixel window [30]. Denoising was by 2D spatial domain Gaussian filtering using the `imgaussfilt` MATLAB™ function [323], with a standard deviation of between 1.8 and 0.45 pixels. Sharpening was by the `imsharpen` MATLAB™ function’s [85] unsharp mask (USM) with a strength of between 1.2 and 2, Gaussian low-pass filter radius of between 3 and 1 pixels, and no thresholding applied.

The non-linear pipeline applied the following algorithms. Demosaicing was by the *One Step Alternating Projections (OSAP)* [70] algorithm. It revises the Alternate Projections [72]

algorithm to perform one filtering step in the polyphase domain [70] to establish inter-channel correlations and mitigate aliasing [72]. The algorithm was set to full convergence. Denoising was by the Block Matching and 3D Filtering (BM3D) [31] algorithm using the ‘normal’ profile. It operates on a patch-wise level as described in Section 2.2.1. It uses the sliding window transform to search for matching image “blocks”. These blocks are assembled and averaged in a three-dimensional (3D) transform domain. Image colour channels were then sharpened separately using the detail enhancement property of the *Guided Image Filter (GIF)* [32]. The filtered channels were then concatenated to obtain the sharpened RGB image. The local window radius of the filter was set to 1. The regularisation parameter was set between 0.02 and 0.045. No external guidance image was specified.

The input parameters of the denoising and sharpening algorithms were tuned for each exposure level to optimise output image quality (judged empirically by the author) after all ISP algorithms were applied. Tuning was carried out on a MacBook Pro Retina (2016) display at a viewing distance of 60cm, giving a Nyquist frequency of 46 cycles/degree.

Data generated at each exposure level is referred to, from here on, by the Poisson noise SNR of the pipeline it was obtained from. However, the SNR of the output images from the pipelines (Equation 2.43) changed after each processing stage and differed from the Poisson noise SNR. For example, prior to denoising, the output image SNR was lower than the Poisson noise SNR, since it also accounts for read noise, dark current noise, and the scaling of Poisson noise in the R and B channels. Denoising and sharpening also affected the output image SNR significantly.

The fifty input images of scenes were selected according to the following rationale:

- 1) The set should represent the subject matter and signal content of commonly captured scenes.
- 2) The set should contain a variety of spatial contents, including naturally occurring and human-made structural signals, textures, edges, and uniform-tone regions.
- 3) Images should be captured under various natural lighting conditions and depths of field.
- 4) Each image must contain minimal artefacts and represent its subject faithfully.
- 5) The minimum horizontal/vertical pixel dimension, d , of each image must satisfy $d > 512$.

Scene images 1-37 and 38-50 are shown in Appendix C and Appendix D, respectively. Scene images 1-17 were from the LIVE Image Quality Assessment Database [225]. Scene images 18-26 and 38-50 were from University of Westminster publications [7], [8], [14], [102]. They were captured using characterised professional digital single-lens reflex (DSLR) cameras. Scene images 27-37 were captured using the same DSLR, lens, focal length, aperture and ISO as scene images from [14].

The dimensions of these images were reduced to 512-by-512 pixels by a combination of bicubic interpolation and cropping according to Equation 4.8. d is the minimum pixel dimension of each image, $F_{IN}(x, y)$. $F_{OUT}(x, y)$ is the resized and cropped image. The functions $B\{\dots\}$ and $C\{\dots\}$ describe bicubic interpolation by a scale factor of 0.5 and cropping to 512-by-512 pixels, respectively. If $d \geq 2048$ the process was repeated until $512 \leq d < 1024$.

$$\begin{aligned} \text{If } d \geq 1024, & \quad F_{OUT}(x, y) = B\{F_{IN}(x, y)\}, & (4.8) \\ \text{elseif } 512 \leq d < 1024, & \quad F_{OUT}(x, y) = C\{F_{IN}(x, y)\}, \\ \text{else} & \quad \text{Reject Image} \end{aligned}$$

Each scene image was then windowed by a bespoke method that tapered its edges to a neutral pixel value but preserved its signal as much as possible (Figure 4.4(b)). This action stopped *periodic replication artefacts*, or wraparound error (present in Figure 4.4(c)) from corrupting its two-dimensional (2D) luminance spectrum (Figure 4.4(d)). These artefacts manifest when the opposite edges of the image of the scene differ in luminance, since the luminance spectrum is computed using the discrete Fourier transform (DFT) that is a periodic function [324, pp. 116–120].

The edges of all scene-and-process-dependent noise images were of similar luminance. Thus, periodic replication artefacts did not corrupt their DFT luminance spectra significantly or affect the accuracy of the various SPD-NPS measures. However, such artefacts biased MTFs measured using unwindowed scene images significantly (as discussed in Section 5.1).

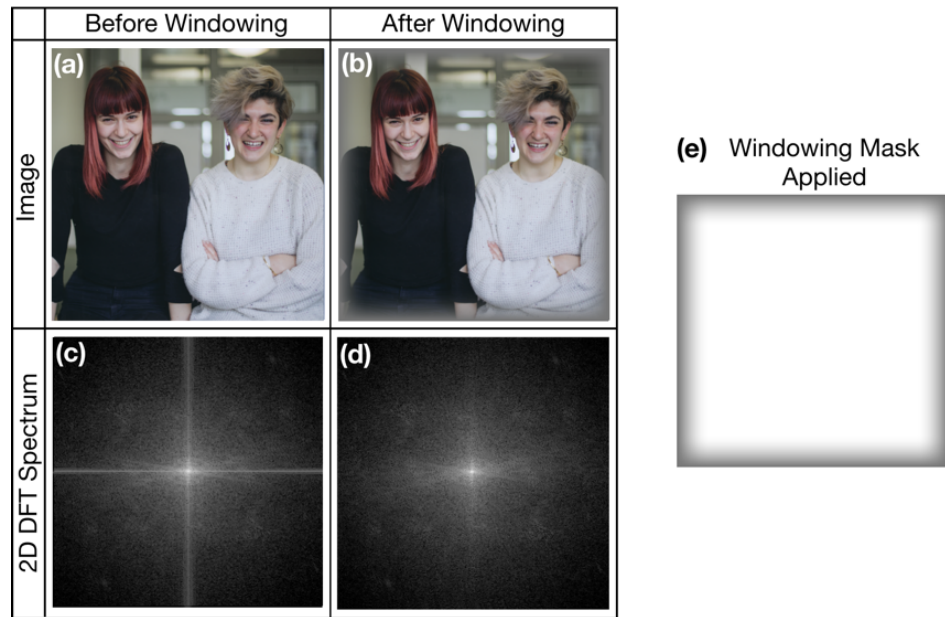


Figure 4.4 Two-dimensional (2D) discrete Fourier transform (DFT) log luminance spectra for the ‘Students’ image [7]: before, (c), and after windowing, (d) using the mask (e).

To window each image, a mask was applied as a transparent layer in Adobe Photoshop™. The mask was generated in MATLAB™ as described below:

- 1) An RGB image of 512-by-512 pixels dimensions and normalised pixel value 1 was created.
- 2) A 2D raised cosine function of 128-by-128 pixels dimensions, frequency $1/128$ cycles/pixel, minimum value 0.46, and maximum value 1 was generated by revising the method of Eddins [325].
- 3) The quadrants of 2) were copied to the corresponding corners of 1).
- 4) The remaining horizontal/vertical edges of 1) were tapered by copying appropriate pixels from the horizontal/vertical edges of the quadrants of 2).

All input images were linearised before being processed by the pipelines. This was achieved using a GOG model that corrected for the gamma of the capture system of 0.4545 (or $1/2.2$); the gain and offset were set to 1 and 0, respectively.

4.3 Results

The various SPD-NPS measures and the uniform patch NPS are presented for the linear and non-linear pipelines at SNRs 40 and 5. These measurements are analysed after each ISP

stage, as indicated by the red arrows in Figure 4.3. Measurements were also taken at SNRs 10 and 20, which showed comparable trends. Each measurement was “smoothed” using a seven-segment moving average filter. MATLABTM code for Burns’ direct dead leaves MTF [29] implementation was adapted significantly to deliver all SPD-NPS measurements, and the uniform patch NPS.

There is no current way to obtain the ground truth (or “correct”) NPS for a given system, due to the effect(s) of: 1) measurement error, 2) the dependency of Poisson noise on input signal intensity, and 3) interactions between the input signal and any non-linear ISP. Thus, the SPD-NPS measures are validated via thorough comparison with existing comparable measures, and expectations of capture system behaviour.

Figure 4.5 compares measurements obtained from dead leaves signals only. It is used to validate the dead leaves SPD-NPS by comparison with the uniform patch NPS. The y-axis of each plot is logarithmically scaled and the test chart images were not windowed, as is common practice in the industry.

Figure 4.6 validates all SPD-NPS measures derived from images of scenes by comparing them with the previously validated dead leaves SPD-NPS measure. Note that, all input images of scenes and test charts were windowed to enable their fair comparison. The y-axes are, again, logarithmically scaled.

Figure 4.7 examines pipeline noise scene-dependency in depth in a linear-linear space. It analyses changes in the distribution and integrated area of pictorial image SPD-NPS measurements across 50 images of different windowed scenes.

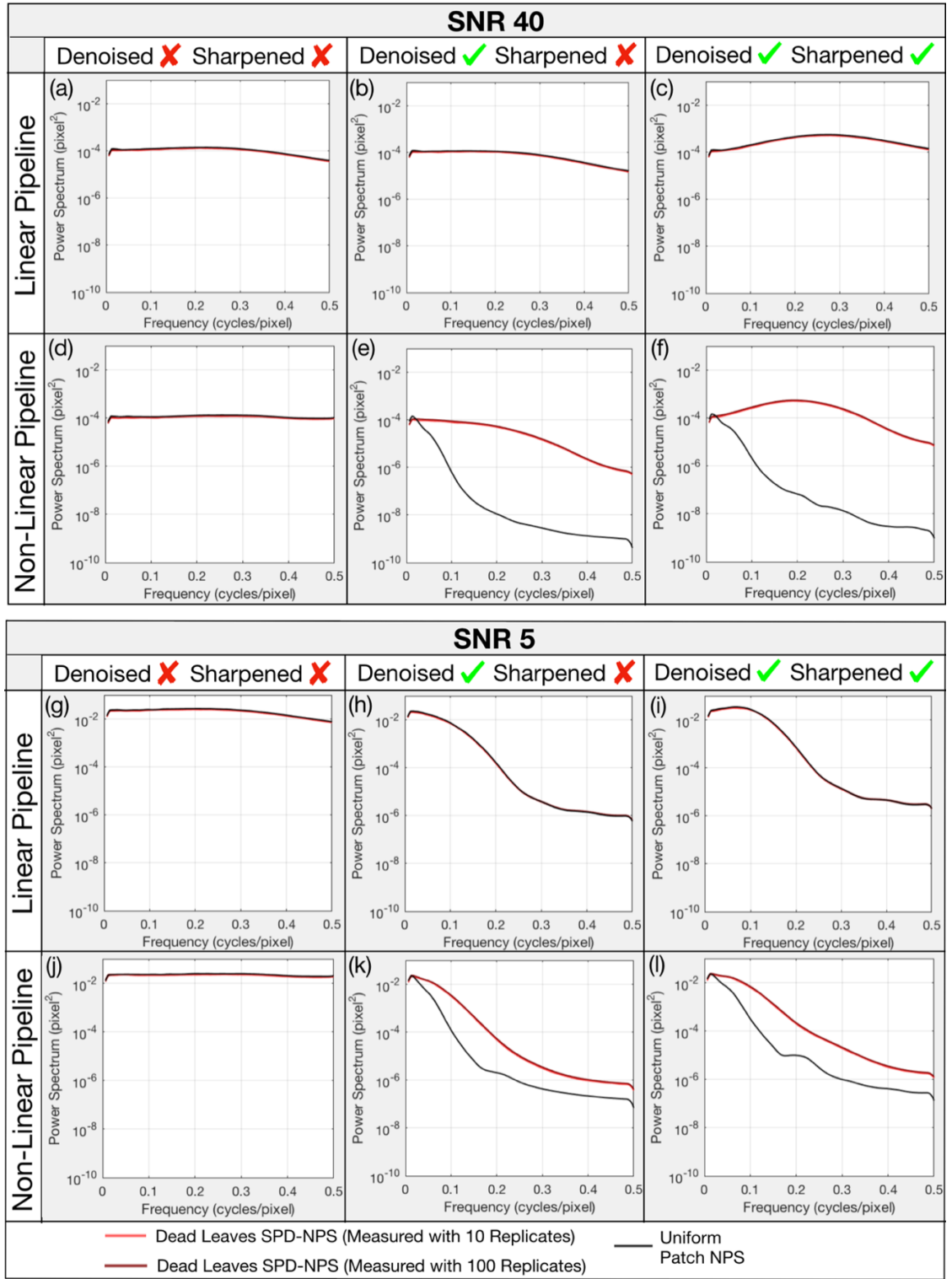


Figure 4.5 Luminance Noise Power Spectra (NPS) and scene-and-process-dependent NPSs (SPD-NPS) derived from different test charts. The uniform patch NPS (grey curves) and proposed dead leaves SPD-NPS (red curves), are shown at different image signal processing (ISP) stages at SNR 40, (a) to (f), and SNR 5, (g) to (l). The power (y) axis is logarithmically scaled.

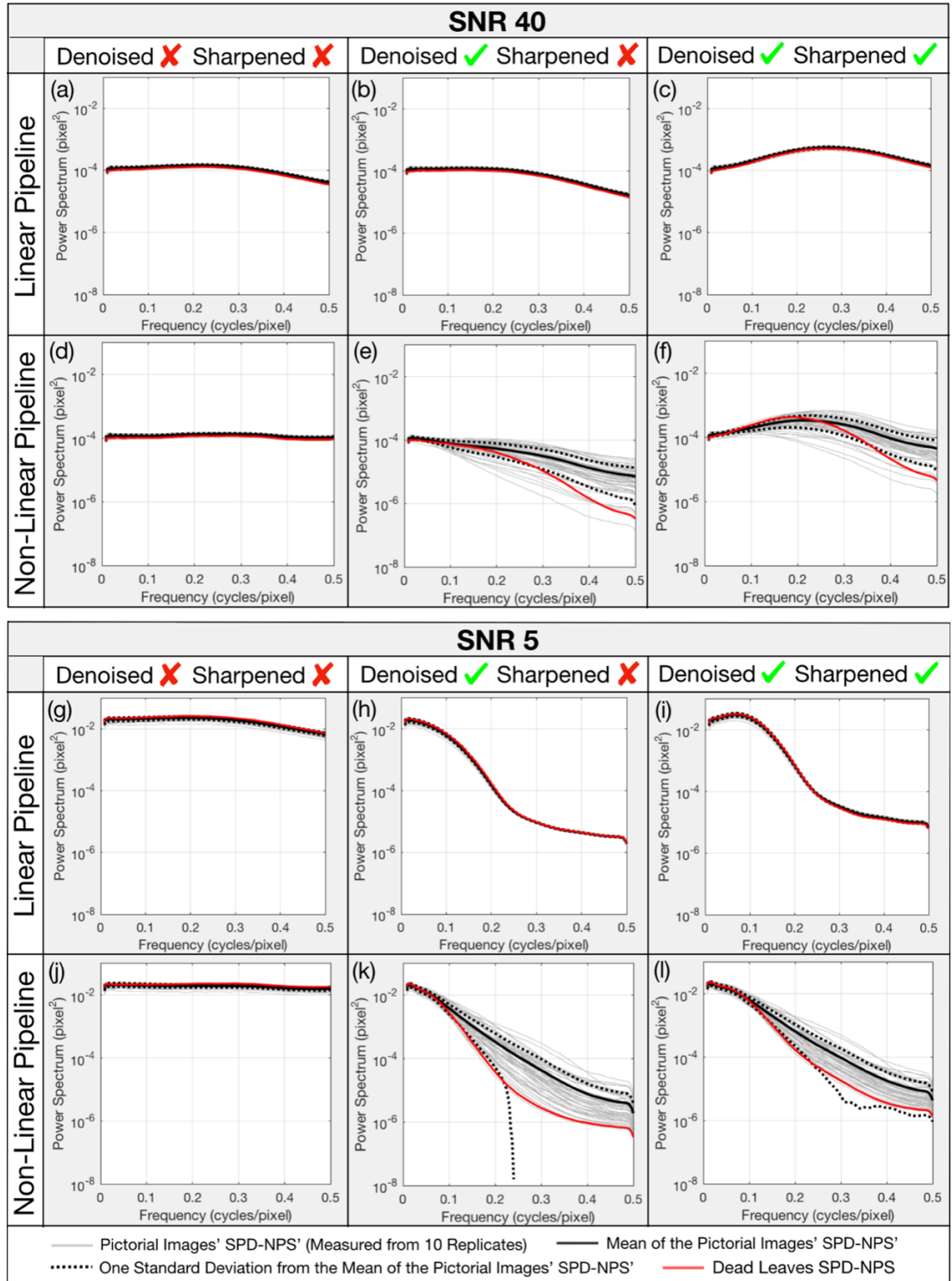


Figure 4.6 Scene-and-process-dependent luminance Noise Power Spectra (SPD-NPS) derived from pictorial images and the dead leaves test chart. The pictorial image SPD-NPS (grey curves), mean pictorial image SPD-NPS (black curves), pictorial image SPD-NPS standard deviation (black dotted curves), and dead leaves SPD-NPS (red curves) are shown after different stages of image signal processing (ISP) at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). The power (y) axis is logarithmically scaled.

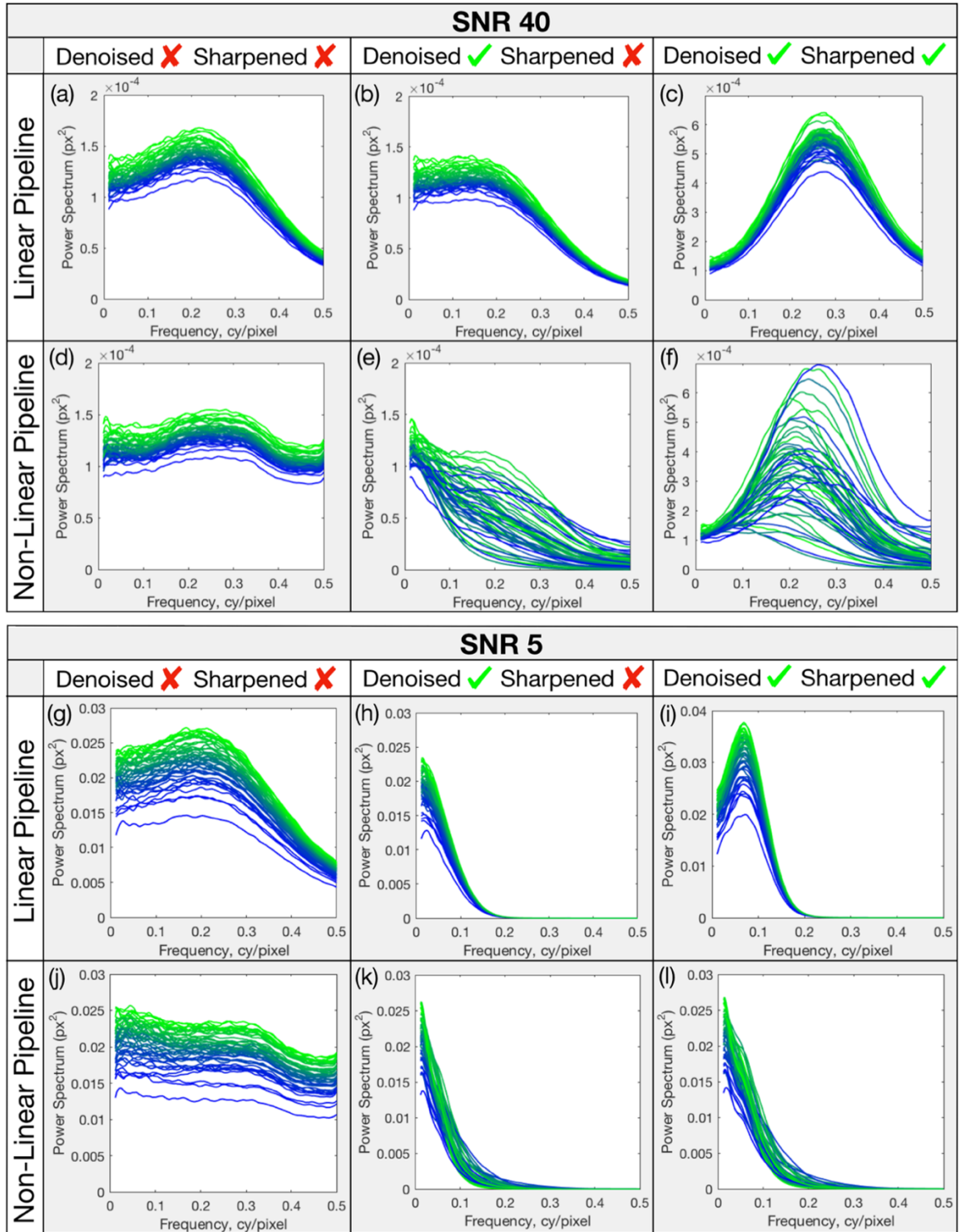


Figure 4.7 Demonstration of luminance noise scene-dependency in the non-linear image capture pipeline. Pictorial image scene-and-process-dependent Noise Power Spectra (SPD-NPS) are plotted on linear axes for the linear and non-linear pipelines at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). Each curve is coloured according to its integrated area before denoising and sharpening. Green curves are of a higher area between zero and Nyquist frequency.

As expected from measures based on linear system theory, the uniform patch NPS was similar to the various SPD-NPS measurements for the linear pipeline. This shows that both the uniform patch NPS and SPD-NPS measures are applicable to the measurement of linear systems.

Figure 4.5 establishes the dead leaves SPD-NPS (red curves) as a more representative (and thus more “accurate”) measure of average real-world noise power for systems using non-linear ISP than the uniform patch NPS (black curves). The uniform patch NPS decreased excessively after non-linear denoising since it presented the denoising algorithm with the simplest possible input signal for denoising. Consequently, after denoising, the uniform patch NPS underestimated the dead leaves SPD-NPS (for which the structure of the dead leaves signal impeded the denoising process). The dead leaves SPD-NPS measurements obtained using 10 and 100 replicates were virtually identical to one another when plotted on logarithmically scaled axes. This validated the computation of further SPD-NPS measures with ten replicates.

Figure 4.6 demonstrates that the pictorial image SPD-NPS (grey curves) accounted suitably for the scene-dependent processing of temporally varying noise by the non-linear pipeline. This is because measurements from images of different input scenes varied significantly only after non-linear content-aware denoising and sharpening algorithms were applied. This observation, coupled with the measure’s limited bias, suggests that implementing it in signal transfer visual IQMs (STV-IQM) would improve correlations with perceived image quality. Likewise, implementing scene-and-process-dependent noise images in the IEEE P1858 Camera Phone Image Quality (CPIQ) metric [22] is also expected to improve the metric’s accuracy.

It was inferred that, the mean pictorial image SPD-NPS (Figure 4.6, black curves) characterised the average real-world level of temporally varying noise in both pipelines effectively. This was due to the apparent lack of bias in the pictorial image SPD-NPS curves, and the fact the scene image set was both large and representative. The pictorial image SPD-NPSs (grey curves) were distributed approximately evenly about the mean pictorial image SPD-NPS (black curves) after all ISP stages, suggesting the latter measure was representative of the “general” performance of the system.

The dead leaves SPD-NPS (Figure 4.6, red curves) underestimated the mean pictorial image SPD-NPS (black curves) after non-linear content-aware denoising and sharpening. This

observation suggests that other content-aware algorithms may also denoise dead leaves signals more effectively than the “average pictorial scene”. The fact the dead leaves chart was denoised differently to images of real scenes was expected, considering the structure of this mathematically generated signal has limited relation to real scenes. Its similarity with scenes is statistical, not physical. It is expected that the non-linear BM3D algorithm denoised the dead leaves signal more effectively for the following reasons. It operates on a patch-wise level (Section 2.2.1). The homogeneity of the dead leaves signal (which consists entirely of discs with perfect edges) allowed the algorithm to find a larger number of “matched” windows than for most images of natural scenes (which contain varied, complex structure).

The scene-dependent variation in the noise in the non-linear pipeline was successfully accounted for by the pictorial image SPD-NPS standard deviation (Figure 4.6, black dotted curves). The measure identified non-linear denoising as the main source of noise scene-dependency, which is valid. Note that, the rapid drop in the lower standard deviation boundary in Figure 4.6(k) is not a discontinuity. It resulted from the measurement crossing the x-axis of a graph with a logarithmically scaled y-axis.

Pipeline noise scene-dependency was examined in detail in Figure 4.7, on linear axes. Pictorial image SPD-NPS measurements from both pipelines showed minor scene-dependent variation before denoising (Figure 4.7(a), (d), (g) and (j)). This is expected to be due to actual scene-dependent variations in: 1) the level of Poisson noise (which is dependent on input signal intensity), 2) the effect of black/white level adjustments on noise, 3) scaling of noise when simulating colour channel quantum efficiency. This variation may also be caused by scene-dependency in the measurement error, which currently cannot be distinguished from genuine system scene-dependency. Regardless of their origin, results from the simulations suggest that such variations do not limit significantly the validity of the pictorial image SPD-NPS. The same applies to the pictorial image SPD-NPS standard deviation and mean pictorial image SPD-NPS that are computed from the latter measure. It should also be noted that these variations in the pictorial image SPD-NPS were not noticeable on logarithmically scaled axes in Figure 4.6, unlike the variations that resulted from application of non-linear content-aware ISP.

The distribution and relative level of spread of the pictorial image SPD-NPS curves was relatively unaffected by the linear ISP algorithms (the curves displayed a smooth transition from green to blue throughout Figure 4.7(a) to (c) and (g) to (i)). However, the shape and

order of the curves changed dramatically after non-linear denoising and sharpening, as well as their relative level of spread. The pictorial image SPD-NPS standard deviation (Figure 4.6, black dotted curves) accounts for changes in their spread. But it does not account for changes in their order which should be considered as a further characteristic of system noise scene-dependency.

4.4 Summary

Three novel SPD-NPS measures were defined. They were measured from replicate captures of relevant input signals, such as scenes. They were validated by analysing measurements from two simulated image capture pipelines that applied either linear or non-linear content-aware ISP algorithms. They accounted for the effect of the input signal on the power of temporally varying luminance noise produced by the pipelines. This included the scene-dependent effects of non-linear ISP.

All measures displayed limited measurement error. For the linear pipeline, they delivered similar measurements to the current standard uniform patch NPS. This suggests they were equally biased to the latter measure. Measurements from the non-linear pipeline displayed significant scene-dependent variation, as would be expected. Thus, all SPD-NPS measures were concluded to be more suitable for non-linear systems than current equivalent measures (if equivalent measures exist). Their limitations are that they do not account for FPN, they are computationally complex, and they require many replicates to be captured and registered.

The pictorial image SPD-NPS was the only measure capable of characterising noise in the non-linear pipeline, with respect to a given input scene image. Likewise, the dead leaves SPD-NPS described the average real-world noise power of this pipeline more appropriately than the uniform patch NPS. The former measure, however, was less representative of average real-world noise than the mean pictorial image SPD-NPS, since non-linear ISP algorithms processed noise in dead leaves signals differently to noise in the average pictorial scene image. A measure for the level of system scene-dependency was also validated successfully but did not account for all relevant scene-dependent behaviour.

Chapter 5 Validation of Scene-and-Process-Dependent MTFs (SPD-MTF)

The *Scene-and-process-dependent Modulation Transfer Function (SPD-MTF) framework* is defined. It characterises the modulation (contrast) transfer of the system versus spatial frequency, with respect to any input signal, accounting for system scene-dependency. In the SPD-MTF derivation, scene-and-process-dependent Noise Power Spectrum (SPD-NPS) measures from Chapter 4 are used to account for system noise. Three novel SPD-MTF measures that implement this framework are presented. Each characterises one of the following:

- 1) system signal transfer with respect to an image of a given input scene;
- 2) the average real-world level of system signal transfer with respect to many images of different input scenes;
- 3) system signal transfer with respect to the dead leaves test chart.

Each measure is validated by analysing measurements from simulated linear and non-linear image capture pipelines, defined in Section 4.2. A novel measure for capture system signal transfer scene-dependency is also introduced and validated.

5.1 Derivation of the SPD-MTF Measures

Traditional Modulation Transfer Functions (MTF) measured from edges, sinusoidal signals and white noise are often unrepresentative of average real-world signal transfer for systems that apply non-linear content-aware image signal processing (ISP) algorithms (as discussed in Section 2.3.2). Average real-world signal transfer in such systems is characterised more suitably by MTFs measured from dead leaves signals that model natural scene statistics (NSS) [16], [17]. It was concluded in Section 2.3.2 that the direct dead leaves MTF implementation [19] (Equation 2.31) is the most appropriate current MTF parameter for image quality metrics (IQM) designed for predicting image quality in such systems.

However, no prior art has verified whether dead leaves signals trigger non-linear content-aware ISP algorithms in the same manner as the “average natural scene”. This assumption must hold true if dead leaves MTFs are to characterise appropriately the average real-world level of signal transfer for systems utilising such processes. It should be noted that Chapter 4 found this assumption did not hold with respect to processing of noise.

MTFs measured from dead leaves signals, or any current test chart for that matter, do not account for scene-dependent and local-content-dependent variations in the signal transfer of non-linear systems. These variations are caused by non-linear content-aware ISP algorithms interacting with pictorial scene signals. It is logical to assume that the accuracy of IQMs and the Noise Equivalent Quanta (NEQ) measure would improve if their MTF parameters accounted for such scene-dependent behaviour. This should reduce the gains and offsets that separate correlations between the IQM output scores and quality ratings for each scene.

The SPD-MTFs of this chapter account for capture system signal transfer scene-dependency. They are based upon the *scene-derived texture MTFs* from Branca *et al.* [7], which extend the direct dead leaves MTF measurement implementation (Equation 2.31) to characterise signal transfer with respect to the one-dimensional (1D) power spectrum, $PS(u)$, of a pictorial image luminance signal, $I(x, y)$. $PS(u)$ is given by the rotational average of $PS(u, v)$ that is defined by Equation 2.30; all other parameters to these equations are as described previously. Branca *et al.* [7] characterised a high-end smartphone camera and a digital single-lens reflex (DSLR) camera by measuring scene-derived texture MTFs with respect to several images of scenes. Scene-dependent variation was higher for the former device, as expected, due to the application of greater levels of non-linear content-aware processing [7].

The resulting MTFs are overestimated, however, for the following two reasons. Firstly, the uniform patch Noise Power Spectrum (NPS) is used to compensate for the biasing effect of system noise. This measure was found in Chapter 4 to be unsuitable for the purpose since it underestimates noise in systems using non-linear content-aware denoising. Thus, the scene-derived texture MTFs [7] are overestimated for low-signal scene images due to signal-to-noise limitations. These limitations are defined by Equation 5.1, where $PS_{Input}(u)$ and $PS_{Output}(u)$ are the input and output scene image power spectra, respectively, and $MTF(u)$ is the scene-derived texture MTF [19]. $NPS_M(u)$ and $NPS_R(u)$ are the measured uniform patch NPS and real system NPS, respectively. These limitations are inherent to the direct

dead leaves MTF measurement implementation; thus, they also apply to the various SPF-MTF measures. It should also be noted that $MTF(u)$ limits to 1 in a theoretical ideal where the real system NPS is measured with absolute accuracy (i.e. $NPS_M(u) = NPS_R(u)$). This is because the numerator and denominator of line 3 of Equation 5.1 limit toward an equal value as $PS_{Input}(u)$ limits toward zero.

$$\text{if } NPS_M(u) < NPS_R(u), \quad (5.1)$$

$$\text{then } \lim_{PS_{Input}(u) \rightarrow 0} PS_{Input}(u) < PS_{Output}(u) - NPS_M(u),$$

$$\text{and } \lim_{PS_{Input}(u) \rightarrow 0} \left(\frac{PS_{Output}(u) - NPS_M(u)}{PS_{Input}(u)} \right) = \infty = MTF(u)$$

Secondly, zero-padding or windowing were not applied to the input scene images before computing the two-dimensional (2D) discrete Fourier transform (DFT) [7]. Thus, periodic replication artefacts were introduced into the luminance spectrum at all frequencies when opposite scene edges differed in luminance; these artefacts are discussed in Chapter 4, Section 4.2.

Simulation of the effect of periodic replication artefacts on the scene-derived texture MTFs are shown in Figure 5.1. They agreed with measurements from real capture systems by Branca *et al.* [7] under comparable conditions. Periodic replication artefacts were unaffected by system processing and thus biased the scene-derived texture MTFs toward $MTF(u) = 1$. The level of bias was scene-dependent. Windowing the input images following the method of Section 4.2 mitigated the artefacts and reduced the bias significantly. However, for the relatively low-power “Contrail” and “Cloud” images, before denoising, removing these artefacts exposed a greater, underlying bias, (shown in Figure 5.1(d)) which originates from signal-to-noise limitations (Equation 5.1).

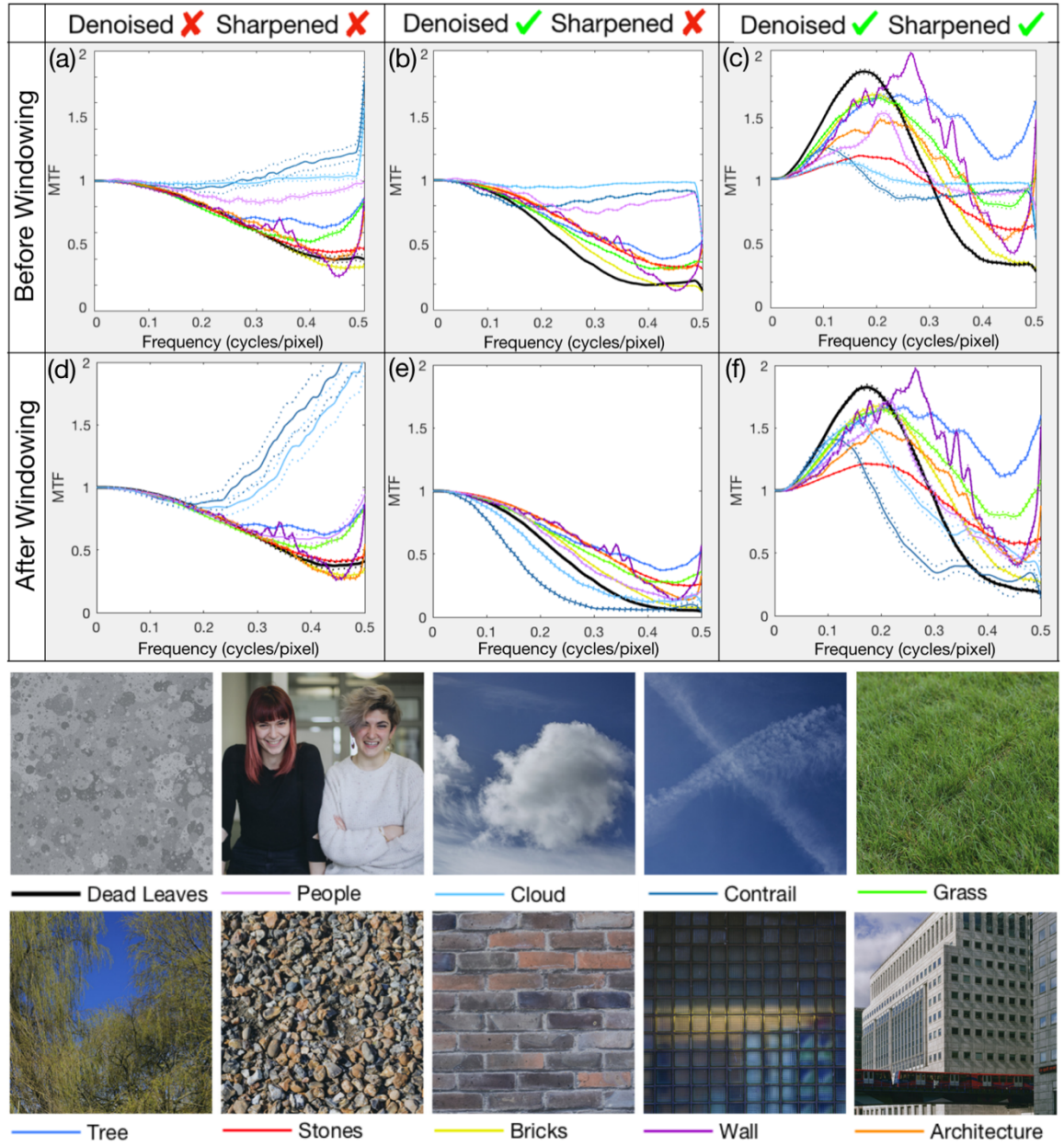


Figure 5.1 Scene-derived texture Modulation Transfer Functions (MTF) of Branca *et al.* [7] for the non-linear pipeline at a signal-to-noise ratio (SNR) of 40, before, (a) to (c), and after the windowing method of Section 4.2 was applied, (d) to (f). Test images are from Branca *et al.* [7], resized to 512 x 512 pixels by bicubic interpolation. Dotted curves show ± 1 standard deviation.

The SPD-MTF framework (Figure 5.2) refines the method of Branca *et al.* [7] to measure the MTF either from windowed images of scenes, or the dead leaves target. It is given by $MTF(u)$ in Equation 2.31, when $PS(u)$ is the rotational average of the 2D DFT luminance power spectrum, $PS(u, v)$, of the scene/target, $I(x, y)$, given by Equation 2.30. $NPS_{output}(u)$ is the SPD-NPS, as measured from the same scene/target (defined in Section 4.1). Utilising SPD-NPS measures instead of the uniform patch NPS mitigates bias from

signal-to-noise limitations (Equation 5.1). All other input parameters to these equations are as stated previously.

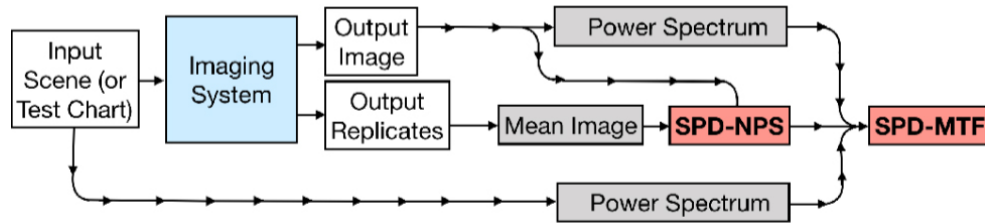


Figure 5.2 The scene-and-process-dependent Modulation Transfer Function (SPD-MTF) framework.

Three SPD-MTF measures are presented in this thesis that implement the SPD-MTF framework. They are defined below and summarised in Table 5.1 alongside the direct dead leaves MTF implementation. They are used as input parameters in various IQMs in Chapter 6. They are more computationally intensive and difficult to implement than current MTFs.

MTF measure	Input parameters	Summary of characterisation	Accounts for system signal transfer scene-dependency?	Accounts for system noise scene-dependency?	Sensitivity to system scene-dependency
Pictorial Image SPD-MTF	Input image PS Output image PS Pictorial image SPD-NPS	Characterises system signal transfer for an individual image	Yes	Yes	Highest ↑ ↓ Lowest
Mean Pictorial Image SPD-MTF	M x input image's PSs M x output image's PSs M x pictorial image's SPD-NPSs	Characterises general system signal transfer	Yes	Yes	
Dead Leaves SPD-MTF	Input dead leaves PS Output dead leaves PS Dead leaves SPD-NPS	Characterises general system signal transfer	Partially	Partially	
Direct Dead Leaves MTF	Input dead leaves PS Output dead leaves PS Uniform patch NPS	Characterises general system signal transfer	Partially	No	

Table 5.1 Summary of the direct dead leaves Modulation Transfer Function (MTF) measurement implementation [19] and the scene-and-process-dependent MTFs (SPD-MTF) of this thesis. PS and NPS are the power spectrum and Noise Power Spectrum, respectively. SPD-NPS is the scene-and-process-dependent NPS. M is the number of images in the test image set.

The *pictorial image SPD-MTF* implements the SPD-MTF framework with respect to a single windowed image of a scene and employs the pictorial image SPD-NPS as the noise measure. It is the only measure capable of characterising the transfer of the signal of a given input scene for systems applying non-linear content-aware processing. It offers improvements over the original measure (i.e. the scene-derived texture MTF [7]) by mitigating bias that results from periodic replication artefacts, as well as the underestimation of noise in systems

applying non-linear denoising. Therefore, scene-dependent variations in the SPD-MTF measurements are more likely to be representative of genuine scene-dependent behaviour of non-linear content-aware ISP algorithms, as opposed to scene-dependent bias. This accuracy will be critical if the measure is to compensate for scene-dependent gain and offset in correlations between IQM scores and psychophysical image quality ratings.

The *dead leaves SPD-MTF* aims to approximate the average real-world signal transfer of the system by passing a dead leaves signal through the SPD-MTF framework. Since the dead leaves SPD-NPS is used to describe system noise, the measure accounts for noise more suitably than the direct dead leaves MTF [19] for systems applying non-linear content-aware ISP. This reduces bias due to signal-to-noise limitations (Equation 5.1). Unlike many pictorial scenes, the dead leaves test chart contains contrast signals at all frequencies that are randomly distributed spatially. Both these characteristics are also expected to help mitigate such bias. However, Section 2.3.2 suggests non-linear content-aware ISP algorithms may not process dead leaves signals as per the “average scene”. This would render the measure unrepresentative of the average real-world level of system signal transfer.

The *mean pictorial image SPD-MTF* is the only measure for the average real-world level of system signal transfer that accounts for processing of real image signals. It is derived as the mean of all pictorial image SPD-MTFs over a large set of images of different scene contents; it is adapted from Jenkin’s MTF-averaging method like the equivalent SPD-NPS measure [2, p. 130]. It is denoted by $N_{Mean_SPD}(u)$ in Equations 4.2 to 4.5 if $N_{SPD}(u)$ is the pictorial image SPD-MTF for the system, $N_{General}(u)$ is the hypothetical “correct” average real-world SPD-MTF for the system, and $\Delta N_{SPD}(u)$ is the difference between them for a single scene image, i , in a set of p scene images. This derivation requires that the scene image set is representative of commonly captured scenes and that $\Delta N_{SPD_i}(u)$ is distributed evenly about zero. The mean pictorial image SPD-MTF is also denoted by $\bar{F}(s)$ in Figure 1.4(b), when $F(s_1)$ to $F(s_n)$ are SPD-MTFs measured from scene images s_1 to s_n , respectively.

Finally, the *pictorial image SPD-MTF standard deviation*, $s(u)$, describes the level of scene-dependent variation in measurements of the signal transfer of a system. Like the equivalent noise measure, this measurement variation can result either from genuine system scene-dependency or scene-dependent measurement error. This unique measure is computed using Equation 4.7, when $N_{Mean_SPD}(u)$ is the mean pictorial image SPD-MTF and $N_{SPD_i}(u)$ is the pictorial image SPD-MTF for each image, i , in a set of p images of different scenes. The

same assumptions and sources of error apply as for the mean pictorial image SPD-MTF. The average real-world level and scene-dependency of signal transfer in a given system is described by subtracting and adding the pictorial image SPD-MTF standard deviation from the mean pictorial image SPD-MTF, as shown by the dotted line in Figure 5.4.

5.2 Results

The SPD-MTFs were validated by the same methodology as the SPD-NPS measures in Chapter 5. Measurements were taken from the simulated linear and non-linear capture pipelines defined in Section 4.2, at signal-to-noise ratios (SNR) of 40 and 5; measurements obtained at SNRs 10 and 20 also displayed comparable trends. All the measurements presented here were smoothed with a seven-segment moving average filter. Burns' MATLABTM implementation [29] was used to calculate the direct dead leaves MTF and was revised significantly to compute all SPD-MTF measures.

It should be noted that, there is no way of deriving the ground truth MTF of a given system for the same reasons as discussed in Chapter 4 for the NPS. Therefore, the various SPD-MTF measures are validated by observing and comparing the measurements from systematic system changes in both pipelines. This method is comparable to that used to validate the SPD-NPS measures in Section 4.3. Figure 5.3 compares the dead leaves SPD-MTF and direct dead leaves MTF, indicating that the former is a more suitable measure for systems applying non-linear content-aware ISP. Figure 5.4 compares the various SPD-MTFs measured from pictorial images with the previously validated dead leaves SPD-MTF. The formatting of this figure is identical to Figure 4.6 that illustrates the equivalent SPD-NPSs. Figure 5.5 applies the same formatting as Figure 4.7 to analyse in detail the scene-dependent characteristics of each pipeline. Each pictorial image SPD-MTF curve is coloured according to its integrated area after demosaicing.

Windowing was not applied to the dead leaves test chart image in Figure 5.3, as is common practice in the industry. But windowing was applied to both the dead leaves test chart and the input scene images in Figures 5.4 and 5.5 to enable fair comparison between measurements obtained from these respective input signals. This is the cause for the minor variations between the dead leaves SPD-MTF measurements in Figures 5.3 and 5.4.

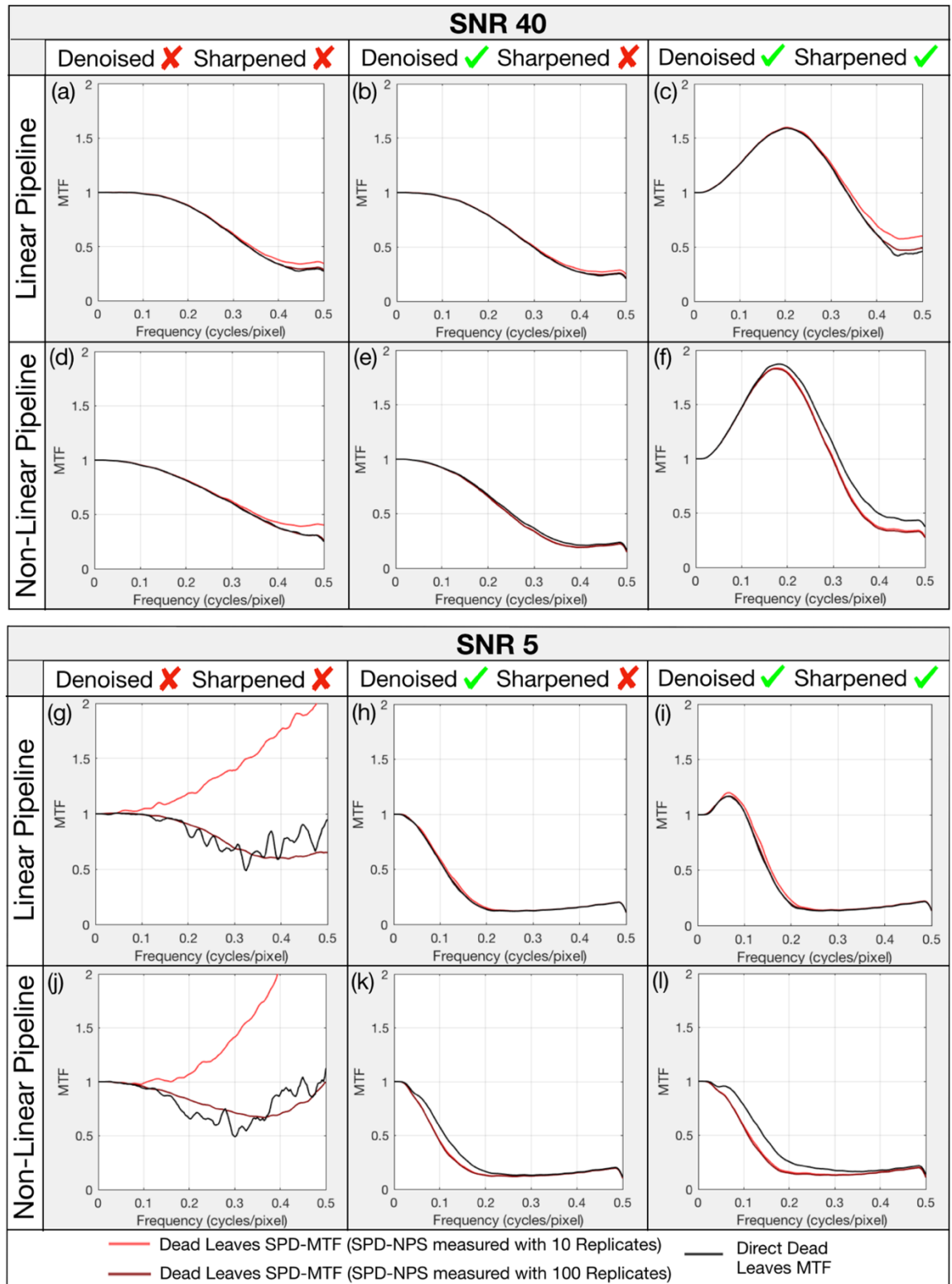


Figure 5.3 Direct dead leaves Modulation Transfer Functions (MTF) (black curves) and proposed dead leaves scene-and-process-dependent MTFs (SPD-MTF) (red curves) after different stages of image signal processing (ISP) at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l).

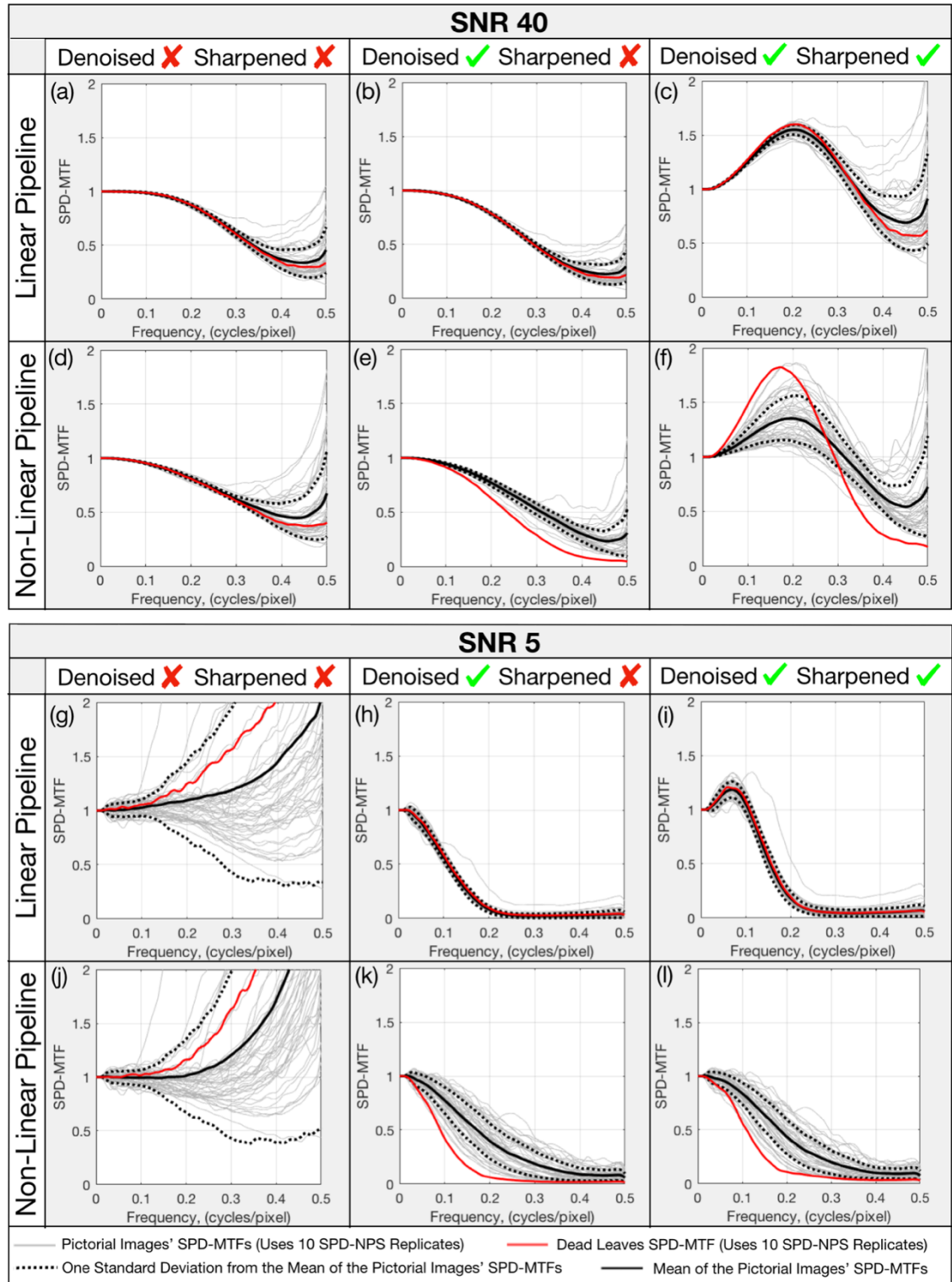


Figure 5.4 Comparison of scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) from pictorial images and the dead leaves test chart. Pictorial image SPD-MTFs (grey curves), mean pictorial image SPD-MTFs (black curves), pictorial image SPD-MTF standard deviations (black dotted curves) and dead leaves SPD-MTFs (red curves) are shown after different image signal processing (ISP) stages at signal-to-noise ratios (SNR) 40, (a) to (f), and 5, (g) to (l).

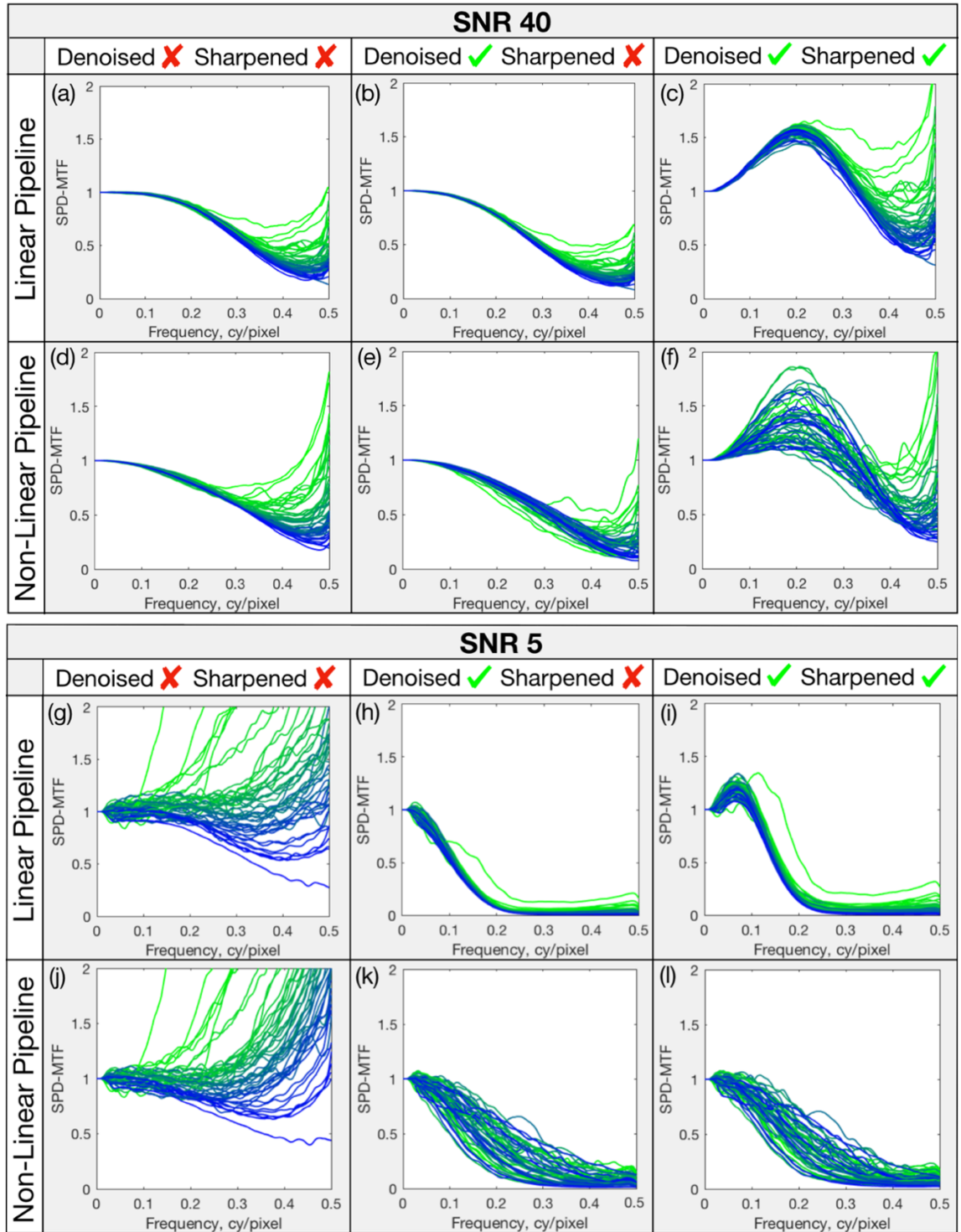


Figure 5.5 Demonstration of signal transfer scene-dependency in the non-linear image capture pipeline. Pictorial image scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) are shown for the linear and non-linear pipelines at signal-to-noise ratios (SNR) of 40, (a) to (f), and 5, (g) to (l). Curves for each scene image are coloured according to their integrated area before denoising and sharpening. Green curves are of a higher area between zero and Nyquist frequency. Ten replicates were used when computing each SPD-MTF.

The experimental conditions at SNR 5 before denoising (i.e. Figures 5.3-5.5 (g) and (j)) were considered less relevant when validating each SPD-MTF measure. This was because it is common for real capture systems to apply denoising at such exposure levels, and the SPD-MTFs were intended for systems that apply denoising and other ISP. All other conditions are referred to as the most relevant conditions.

All measures were heavily biased under these less relevant conditions due to signal-to-noise limitations expressed in Equation 5.1. Denoising mitigated this bias since it reduced the absolute difference between the measured, $NPS_M(u)$, and real noise power, $NPS_R(u)$. Figure 5.3 (g) and (j) demonstrate how raising the number of SPD-NPS replicates reduces this bias by increasing the accuracy of the SPD-NPS parameter, $NPS_M(u)$.

Under the most relevant conditions, the dead leaves SPD-MTFs (Figure 5.3, red curves) and direct dead leaves MTFs (black curves) were of similar shape for the linear pipeline, as expected from linear system theory. Measurement bias was marginally lower when the dead leaves SPD-MTF was measured using 100 replicates, compared to 10 replicates. The similarity between these measurements justified the decision to compute other SPD-MTF measures shown in Figures 5.4 and 5.5 with 10 replicates.

Figure 5.3(k) and (l) show that the direct dead leaves MTF was overestimated moderately after non-linear denoising at SNR 5. Figure 4.5(k) demonstrates that this bias was due to signal-to-noise limitations (Equation 5.1) since the uniform patch NPS, $NPS_M(u)$, underestimated the real noise power, $NPS_R(u)$, by a significant margin under these conditions. Sharpening compounded this bias at SNR 5 (Figure 5.3(l)) and also rendered it noticeable at SNR 40 (Figure 5.3(f)). Thus, the dead leaves SPD-MTF was established as the more suitable measure for signal transfer in non-linear systems under the most relevant conditions.

After denoising, the scene-dependent variations between pictorial image SPD-MTFs from the non-linear pipeline were far higher than for the linear pipeline, as demonstrated in Figures 5.4 and 5.5. This important observation is a direct result of the measure accounting for interactions between the scene image content and the pipeline's non-linear content-aware ISP algorithms, with respect to both signal transfer and noise.

However, pictorial image SPD-MTFs from the linear pipeline also displayed significant scene-dependent variation under some of the most relevant conditions (Figures 5.4(a) to (c),

grey curves). It is not currently possible to diagnose the source of this particular scene-dependent variation. It is expected to be partly caused by signal transfer in the linear pipeline being slightly scene-dependent, resulting from interactions between scene image signals and processing during the modelling of Poisson noise, black/white level adjustments and colour channel quantum efficiency. The primary cause for the scene-dependent variation, however, is expected to be scene-dependent bias from signal-to-noise limitations (Equation 5.1). This bias affected particularly the higher frequencies of the low-signal images of scenes, at lower SNRs, where the signal power, $PS_{Input}(u)$, was closer to zero and the pictorial image SPD-NPS, $NPS_M(u)$, was more likely to underestimate the real noise power, $NPS_R(u)$. It should be noted that the scene-dependent variation was much lower after linear denoising, especially at lower SNRs. This is attributed to the bias being mitigated, since denoising reduced the absolute difference between the measured noise power, $NPS_M(u)$, and real noise power, $NPS_R(u)$, more than it reduced the output signal power, $PS_{Input}(u)$.

The pictorial image SPD-MTF is the most theoretically valid MTF measure in situations where the signal transfer of a non-linear system must be measured with respect to a given input scene (e.g. in certain image quality modelling applications). Deriving the MTF in this way, however, increased measurement error for the reasons explained above, which affected the measure's accuracy. The trade-offs between these advantages and disadvantages of the measure are scene-dependent. They need to be investigated from first principles in further work, for example, by adapting the direct dead leaves MTF error propagation method of Burns [158]. Performing such an investigation was outside the scope of this thesis. Instead, in Chapter 6 of this thesis, the validity of the pictorial image SPD-MTF is evaluated further, experimentally, by assessing whether implementing the measure improves the accuracy of various IQMs.

The mean pictorial image SPD-MTF (Figure 5.4, black curves) and dead leaves SPD-MTF (red curves) had similar shape for the linear pipeline under the most relevant conditions. It can therefore be inferred that, the average level of measurement error across the pictorial image SPD-MTFs from the 50 input images was approximately equal to the error in the dead leaves SPD-MTF. It can also be inferred that the scene-dependent nature of the bias in the former caused the bias to “average out” across the 50 measurements taken. This would suggest that, provided that the mean pictorial image SPD-MTF is measured from a representative set of images of different scenes, it is not significantly affected by the bias in the individual pictorial image SPD-MTF measurements from which it is derived.

After non-linear denoising and sharpening, however, the dead leaves SPD-MTF was an outlier compared to the pictorial image SPD-MTFs and generally underestimated the mean pictorial image SPD-MTF. Observations in the previous paragraph suggest these differences were not due to measurement error. It is concluded that the signal transfer of the non-linear pipeline's ISP algorithms was different for dead leaves signals than for the "average scene".

More specifically, dead leaves signal transfer was reduced more by non-linear denoising than signal transfer for the average pictorial scene image (Figure 5.4(e) and (k)). The global power spectrum of natural scenes follows a $1/f$ power law with respect to spatial frequency, like the dead leaves chart [16], [17]. Their contrast signals are often clustered spatially, however, not randomly distributed as per the dead leaves chart. Thus, edges in the average natural scene would be expected to have, on average, higher contrast than edges in the dead leaves target. The main side-effect of non-linear denoising is removal of texture (i.e. low-contrast fine details). It is, therefore, understandable that the signal transfer of the lower-contrast edges of the dead leaves target was particularly affected by it.

The signal transfer of dead leaves signals was also boosted more by non-linear sharpening than the average pictorial scene (Figure 5.4(f)). This is expected to be due to the edges of all circles in the dead leaves signal being "perfect" (i.e. with maximum gradient). Perfect edges are more responsive to non-linear sharpening algorithms than edges of lower gradient [146].

The mean pictorial image SPD-MTF is expected to be a more suitable measure than the dead leaves SPD-MTF for the average real-world signal transfer of non-linear systems. This was despite the former measure inheriting some bias from the pictorial image SPD-MTFs. It could not be concluded whether $\Delta N_{SPD_i}(u)$ in Equations 4.2 to 4.5 was distributed evenly about zero for the pictorial image SPD-MTFs. However, the distribution of the pictorial image SPD-MTFs around the mean pictorial image SPD-MTF did not appear skewed under the most relevant conditions.

The pictorial image SPD-MTF standard deviation (Figure 5.4, black dotted curves) indicated system scene-dependency was significantly higher after application of the non-linear ISP algorithms, than the equivalent linear algorithms. This was expected to be because of the increased algorithm content awareness; the difference was particularly clear after intense denoising at low SNRs. The value of this scene-dependency measure was, however, larger for the linear pipeline than as expected, and as demonstrated in practice by the equivalent SPD-NPS measure (Figure 4.6(a) to (c) and (g) to (i), black dotted curves). This was caused

by scene-dependent variation in the pictorial image SPD-MTF measurements, which was mainly due to bias from signal-to-noise limitations. It should also be noted that the order of the pictorial image SPD-MTF curves changed significantly after non-linear ISP algorithms were applied (Figure 5.5). The pictorial image SPD-MTF standard deviation did not account for this.

The integrated area under each pictorial image SPD-MTF curve (Figure 5.5) correlated relatively well with the *busyness* [59] of each scene image, before denoising. Applying non-linear ISP algorithms disturbed this relationship significantly. The busyness descriptor expresses, as a single figure, the proportion of the image containing higher frequencies exceeding a given contrast threshold [59]. Thus, it accounts for the spatial distribution and power of higher-frequency signals, which are both relevant to the signal-to-noise limitations that bias the SPD-MTFs.

Linear Pipeline					
	SNR 40	SNR 20	SNR 10	SNR 5	Mean of all SNRs
Denoised ✗ Sharpened ✗	0.559	0.690	0.758	0.765	0.693
Denoised ✓ Sharpened ✗	0.560	0.651	0.474	0.547	0.558
Denoised ✓ Sharpened ✓	0.591	0.636	0.292	0.506	0.506
Non-Linear Pipeline					
	SNR 40	SNR 20	SNR 10	SNR 5	Mean of all SNRs
Denoised ✗ Sharpened ✗	0.438	0.678	0.763	0.768	0.662
Denoised ✓ Sharpened ✗	0.231	0.196	0.074	0.013	0.129
Denoised ✓ Sharpened ✓	0.043	0.003	0.043	0.010	0.025

Table 5.2 R^2 correlation coefficients of a logarithmic curve fit of form $y = m \cdot \ln(x) + c$, to the regression between the integrated area under the pictorial images' scene-and-process-dependent Modulation Transfer Functions (SPD-MTF) and the busyness' of these scenes [59].

5.3 Summary

Three novel SPD-MTF measures were proposed. They characterised system signal transfer with respect to relevant input signals, accounting for the scene-dependent effect of non-linear ISP algorithms. Further, a measure describing the level of system scene-dependency was presented. All measures were validated by analysing measurements from simulated linear and non-linear camera pipelines.

The dead leaves SPD-MTF offered minor improvements over the current direct dead leaves MTF measurement implementation [19]. But the signal transfer characteristics of non-linear content-aware ISP algorithms were different for dead leaves signals compared to the average pictorial scene image. This may affect the relevance of both these measures to capture system design and image quality modelling.

The pictorial image SPD-MTF accounted most comprehensively for system scene-dependency. But it suffered from scene-dependent bias due to signal-to-noise limitations, despite attempts to mitigate it. These limitations were inherent to the direct dead leaves MTF, which the proposed measures are based upon. The resultant bias was difficult to distinguish from genuine effects of system scene-dependency. It affected the accuracy of the pictorial image SPD-MTF standard deviation. Nevertheless, it averaged out across the 50 test scenes to levels comparable to bias in the direct dead leaves MTF. Thus, the mean pictorial image SPD-MTF is concluded to describe average real-world system signal transfer appropriately, but it is computationally inefficient when compared to both the dead leaves SPD-MTF and direct dead leaves MTF.

Chapter 6 An Evaluation of Scene-and-Process-Dependent IQMs

A number of variations of five spatial image quality metrics (IQM) are evaluated in this chapter, referred to as *variants*. Two of these IQMs are novel (log Noise Equivalent Quanta (log NEQ) and Visual log NEQ) and are based on the novel scene-and-process-dependent Noise Equivalent Quanta (SPD-NEQ) performance measure. The other IQMs are scene-and-process-dependent versions of the Square Root Integral with Noise (SQRI_n) [12], Perceived Information Capacity (PIC) [169] and IEEE P1858 Camera Phone Image Quality (CPIQ) metrics [22]. Each of these metrics is defined in Chapter 3.

6.1 Methodology

Variants of each IQM were generated in MATLABTM (Section 6.1.3). Each variant used a different permutation of input parameters for imaging system noise (Table 4.1), signal transfer (Table 5.1) and human visual system (HVS) sensitivity (Table 6.1). Thus, each variant accounted for imaging system and visual scene-dependency to a different degree.


Contrast Sensitivity Function (CSF)	Input Parameters	Scene-dependent Visual Behaviour Accounted For	Sensitivity to Visual Scene-Dependency
Johnson & Fairchild CSF	• N/A	• N/A	Lowest  Highest
Barten CSF	<ul style="list-style-type: none"> • Mean luminance of the pictorial scene • Luminance of the interface's background 	• Adaptations with respect to global image luminance	
Contextual CSF (cCSF) / Contextual Visual Perception Function (cVPF)	<ul style="list-style-type: none"> • The pictorial scene's luminance contrast spectrum • Luminance of the interface's background 	<ul style="list-style-type: none"> • Adaptations with respect to global image luminance • Contrast masking 	

Table 6.1 Summary of the contrast sensitivity functions (CSF) employed in Chapter 6.

Output scores for each IQM variant were recorded over a series of test images that were generated following the method presented in Section 6.1.1. Twenty-seven observers rated the quality of the test images using the ISO 20462 [56] softcopy image quality ruler psychophysical paradigm (Section 6.1.2). These ratings were expressed on a Standard Quality Scale (SQS₂) scale [56, p. 3], a ratio scale with calibrated intervals of 1 just-noticeable difference (JND) in quality ranging from SQS₂ 0 to 31. The zero-point describes

very low-quality pictorial images, the subjects of which are difficult to identify. Thirty one represents the highest quality.

The IQM variants were evaluated by benchmarking their accuracy and level of correlation with respect to the observer quality ratings in Section 6.2.2. The behaviour of selected IQM variants is analysed further in Section 6.2.3. These evaluations informed upon which input parameters were most appropriate for image quality modelling and which IQMs responded most favourably to modification.

6.1.1 Test Image Dataset

The test image dataset was generated using simulated image capture pipelines implementing linear and non-linear image signal processing (ISP) algorithms, presented in Section 4.2. The input pictorial images to these pipelines were created from 14 original images that were captured and processed by Allen [102] using the method described in Appendix E. According to reference [102] their quality in terms of SQS_2 was 23 (relatively high). Each image was resized to 512-by-512 pixels dimensions according to Equation 4.8 before being input into the pipelines.

The input images to the simulations are shown in Appendix D and were selected to prioritise the following:

- 1) heterogeneity of scene subjects;
- 2) diversity of objective signal contents, including natural and human-made structural signals, colours, textures, and smooth tonal gradations;
- 3) variation of scene capture settings, including focus distance, focal length, and depth of field;
- 4) introduction of simulation pipeline artefacts that:
 - a. were representative of real capture systems according to the author;
 - b. were not overly spatially concentrated.

Point 4)b encouraged observers to base their opinions on artefacts across the whole image. This was a logical choice since the IQM input parameters were global measures.

The simulation pipelines were identical to those employed to validate the scene-and-process-dependent Noise Power Spectrum (SPD-NPS) and equivalent Modulation Transfer Function (SPD-MTF) measures in Chapters 4 and 5, respectively; they are described in Section 4.2.

The following modifications were applied to the pipelines:

- 1) The opacity of the linear and non-linear ISP filters was optimised according to Table 6.2, to maximise the perceived image quality under the viewing conditions, after combined denoising and sharpening.
- 2) Two-dimensional (2D) photon noise was simulated at maximum linear signal-to-noise ratios (SNR) of 10, 20, 40 and 80, according to Equation 2.32.
- 3) Images with noise simulated at SNR 5 were omitted; pilot experiments showed that they were often outside the range of the SQS_2 values.

Appendix I presents all SPD-NPS and SPD-MTF measures, as well as the uniform patch NPS and direct dead leaves Modulation Transfer Function (MTF) for these pipelines.

Pipeline Type	ISP Type	Filter	Opacity			
			SNR 10	SNR 20	SNR 40	SNR 80
Linear	Denoising	Gaussian	85%	83%	82%	80%
	Sharpening	USM	60%	60%	55%	55%
Non-Linear	Denoising	BM3D	87%	86%	86%	85%
	Sharpening	GIF	60%	70%	65%	60%

Table 6.2 Optimal opacities for the pipelines' Image Signal Processing (ISP) filters described in Section 4.2. BM3D, GIF, USM and Gaussian refer to the Block Matching with 3D Filtering [31], Guided Image Filter [32], MATLABTM `imsharpen` unsharp mask [85] and Gaussian blur filters, respectively.

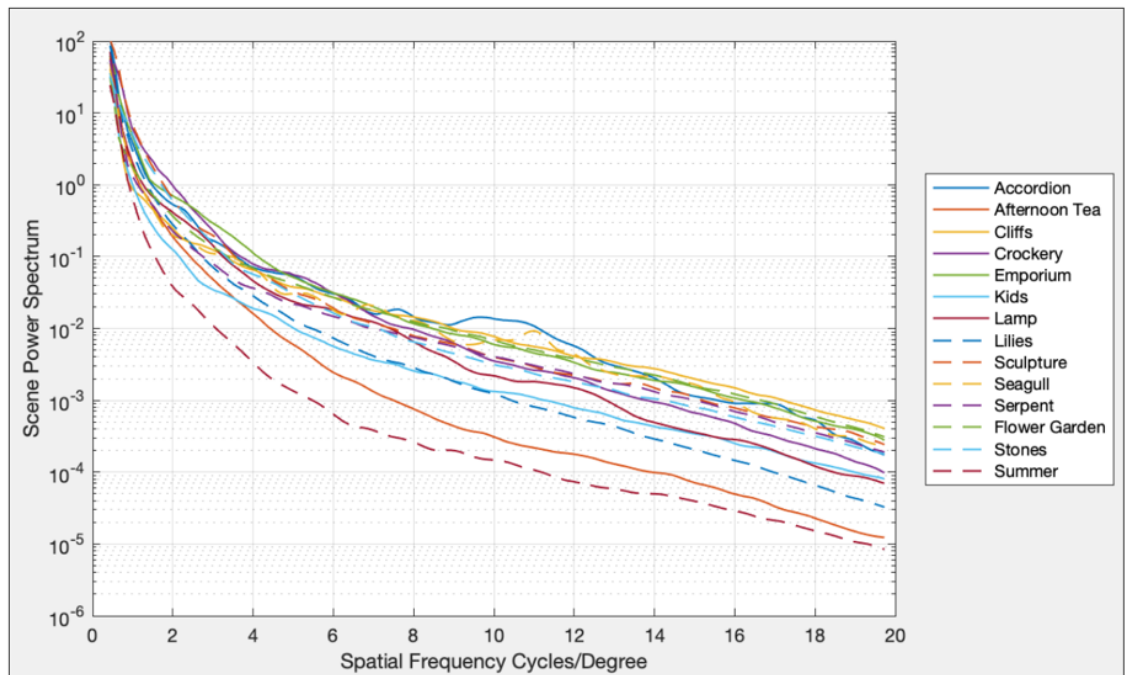


Figure 6.1 Power spectra for the input images to the simulations (shown in Appendix D).

According to Equation 6.1, reducing the percentage opacity (P) of the ISP filters increased their transparency by blending a proportion of the image before filtration, $g(x, y)$, with a proportion of the filtered image, $d(x, y)$. This improved subjective image quality by reducing ISP filter artefacts in the output image, $o(x, y)$, at the expense of slightly higher noise which helped to mask the artefacts. Lowering the opacity of the ISP filters provided greater challenges for testing the robustness of the SPD-MTF and SPD-NPS algorithms, since these measures were designed to account for filtered image signals and noise. It was also the only method of lowering the intensity of certain ISP filters to perceptually optimal levels (tuning), at certain SNRs. Appendix I discusses the effect of reducing the opacity of the ISP filters on all SPD-MTFs and SPD-NPSs.

$$o(x, y) = \frac{P}{100} \cdot d(x, y) + \frac{100 - P}{100} \cdot g(x, y) \quad (6.1)$$

Test images from each pipeline were saved as lossless Portable Network Graphics (PNG) [326] files after demosaicing, denoising and sharpening ISP stages, shown by red arrows in Figure 4.3. Fifty-six images were output from each pipeline after each ISP stage, covering all permutations of the 14 input original images at 4 SNRs. Images from the linear pipeline were chosen to represent both pipelines before denoising was applied. This was because the linear and non-linear demosaicing algorithms produced very similar artefacts to one another when they were evaluated subjectively; they also affected the system performance measurements to a similar degree. A total of 280 test images were generated.

6.1.2 Psychophysical Evaluation

The aim of the psychophysical evaluations was to record subjective quality ratings for the test image dataset. The ISO 20462 [56] softcopy image quality ruler was employed for its speed and accuracy when measuring image quality differences spanning over many JNDs. It has been validated [327] and is implemented widely. The perceived quality of each test image was rated by matching its quality with an image from a set of ordered, univariate reference stimuli differing in sharpness by intervals of 1 JND in quality. Allen [102, p. 216] generated these “ruler images” as described in Appendix E, following the recommendations of the ISO 20462 standard [56, p. 9]. The graphics user interface (GUI) employed for the evaluations was also developed by Allen following this standard, shown in Figure 6.2. The GUI background was set to a uniform neutral tone of luminance 26.6 cd/m^2 . Test and ruler

images were shown on the right and left-hand sides of the display, respectively. Test images were presented in a randomised order. The ruler images were of the same scene as the test image. The observer adjusted the ruler image quality to match the quality of the test image using the keyboard's arrow keys, or by moving the slider. Once they judged both images to be of equal quality, they pressed the "Next Image" button. This recorded the SQS₂ value of the ruler image, presented the next test image, and randomised the position of the slider.

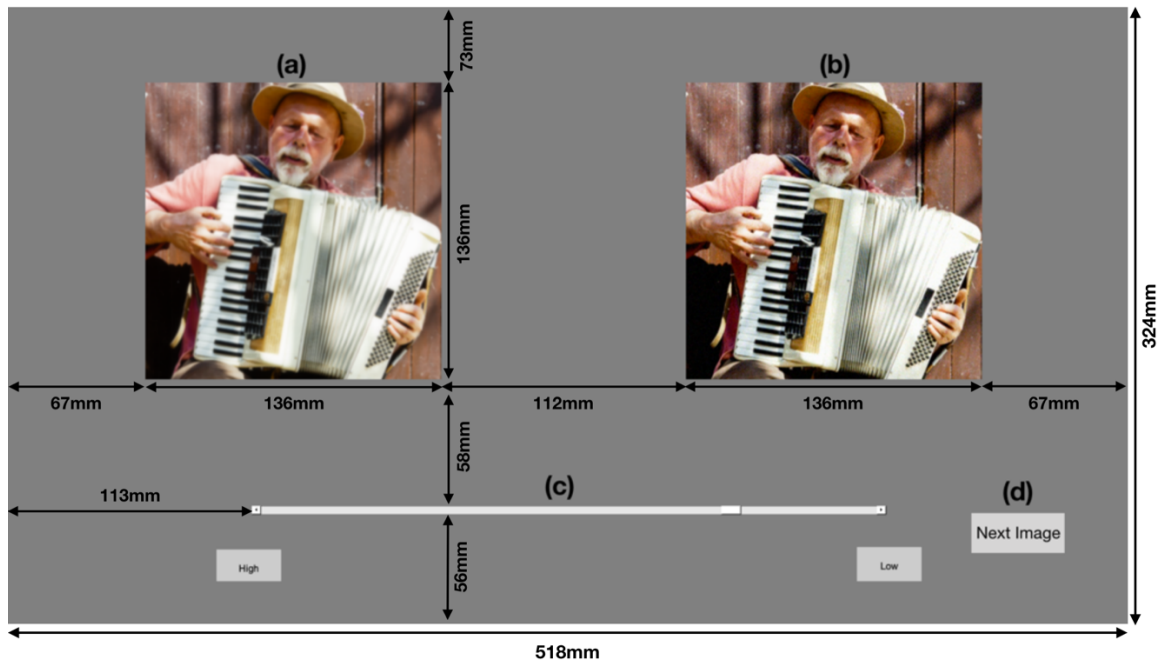


Figure 6.2 Layout of the image quality ruler graphics user interface (GUI) [102, Sec. 6.5.2]: (a) ruler image, (b) test image, (c) slider to select ruler images, (d) button to select next test image.

Figure 6.3 shows the layout of the laboratory equipment. The EIZO ColorEdge CG245W [328] liquid crystal display (LCD) was calibrated to the standardised RGB (sRGB) colour space and characterised as described in Appendix F. The viewing distance of 60cm was restricted using a head-rest, giving a display Nyquist frequency of 20 cycles/degree. The horizontal and vertical viewing angles of the GUI were 47.7 and 30.2 degrees, respectively. Other viewing conditions remained constant and close to the typical office viewing environment described in the sRGB standard [329]. The luminance of the background, desk and table were 25.84 cd/m², 26.4 cd/m² and 41.31 cd/m², respectively, under the moderate ambient illuminance level of 106.4 lux, from a light source with Correlated Colour Temperature (CCT) of approximately 3600K (with CIE x and y chromaticities of 0.4083 and 0.4126, respectively).

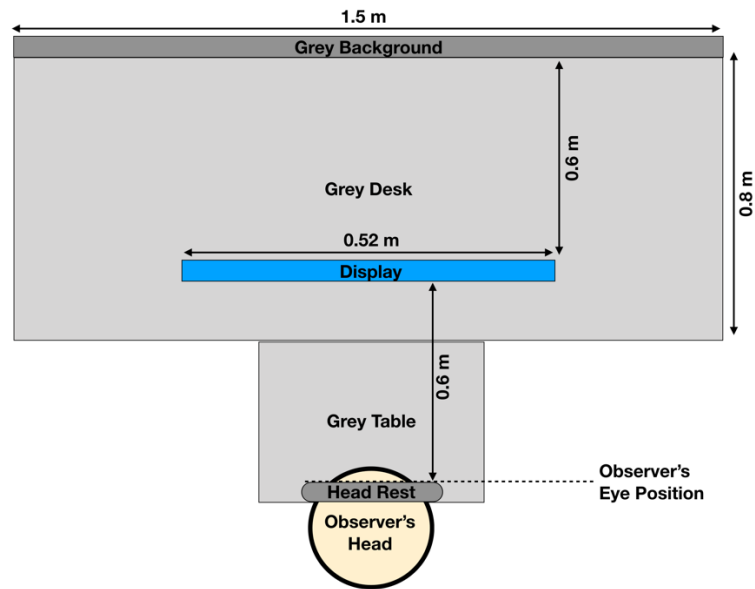


Figure 6.3 Layout of laboratory equipment (plan view).

The modelled display MTF [22, p. 16] used by each IQM (Equation 3.2) predicted Allen's [102, p. 212] previous display MTF measurements adequately, as shown in Figure 6.4. The camera-lens-display system reproduced all visible frequencies at the viewing distance.

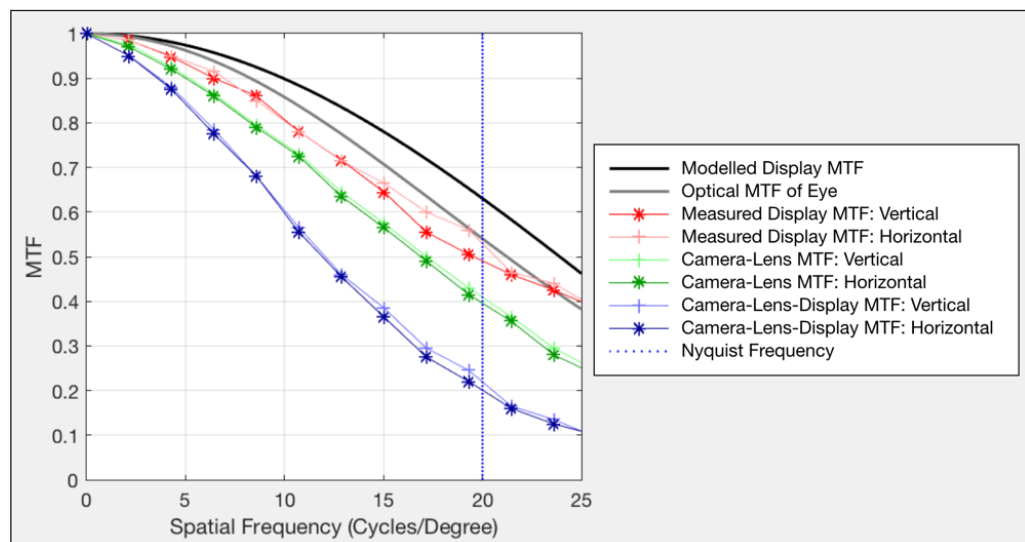


Figure 6.4 Modulation Transfer Functions (MTF) of imaging chain components and Barten's optical MTF of the eye [183, p. 29]. The camera-lens, camera-lens-display and display MTFs were all measured by Allen [102, p. 212]. The latter was also modelled by Equation 3.2 [22, p. 16].

The experimental conditions were very close to the recommendations of ISO 20462, Clause 6.1, for implementing the softcopy ruler, as indicated in Table 6.3 [56, Sec. 6.1].

	Conditions Recommended by ISO 20462 (Clause 6.1)	Conditions of Experiment
Maximum Observer-to- Display Viewing Distance Variation:	4% of Average Viewing Distance	2% of Average Viewing Distance
Maximum Variation in Viewing Distance Across the Stimuli	2% of Arithmetic Average Viewing Distance	2.33% of Arithmetic Average Viewing Distance
Maximum Variation in the Viewing Angle from Perpendicular (at the Centre of Each Stimulus)	10 degrees	12.6 degrees
Recommended Minimum Viewing Distance (2500 x pixel pitch)	0.675 m	0.6 m

Table 6.3 Comparison of the conditions of the psychophysical image quality evaluations with the recommendations of ISO 20462 [56, Sec. 6.1].

The observers wore corrective spectacles/lenses if required for the viewing distance. Before participating in the evaluations, each observer tested their visual acuity, under supervision, by attempting to read row 7 of the Snellen near vision test card [330] (shown in Appendix G) under the viewing conditions. Two prospective observers failed and did not participate since their spatial visual sensitivity did not reflect the contrast sensitivity function (CSF) models used by the IQMs. Two males passed this acuity test but suffered from red-green colour blindness. They were allowed to participate because such observers provided similar responses to trichromats in a comparable spatial image quality study [48, p. 59]. Twenty-seven observers qualified for the evaluations, including 17 males and 10 females of various ethnicities, with an approximate age range of 20 to 55. Six observers had previous experience evaluating image quality or attribute strength. They are referred to herein as experienced observers. All others are referred to as inexperienced observers.

The evaluations were divided into three stages, according to the three ISP stages shown in Figure 4.3. Before participating, each observer read the written instructions shown in Appendix H that included examples of images generated after each ISP stage. They were then trained in a short trial run of the experiment, where they were encouraged to explain their decision making and ask questions. They took between 20 and 45 minutes to complete each stage, taking short breaks in between to avoid fatigue.

6.1.3 Generation of the IQM Variants

A total of 332 IQM variants were generated. They included variants of the log NEQ and Visual log NEQ as well as scene-and-process-dependent versions of the CPIQ metric and the PIC and SQRIn. Each variant employed a different permutation of the NPS, MTF and

CSF parameters listed in Tables 4.1, 5.1 and 6.1, respectively. Appendix I presents and discusses all input parameters. Note that the various NPS and MTF measurements differ from those of Chapters 4 and 5, respectively, since the pipelines' ISP filters were tuned at reduced opacity, and images were generated at SNRs 10 to 80.

The MATLABTM code for Burns' direct dead leaves MTF implementation [29] was adapted to compute all SPD-NPSs and SPD-MTFs, as in Chapters 4 and 5. The CPIQ visual noise metric was computed using MATLABTM code [331] from Baxter and Murray's [332] implementation.

When the Barten CSF [186], Contextual CSF (cCSF) [35] and Contextual Visual Perception Function (cVPF) [8] were employed in the CPIQ metric, they were normalised to the same integrated area as Johnson and Fairchild's [188] luminance CSF from the IEEE P1858 standard [22, p. 72], unless otherwise stated. Likewise, the cCSF/cVPF were normalised to the same integrated area as Barten's CSF when they were employed by the PIC, SQRIn, or Visual log NEQ, unless stated otherwise. In both cases, this stopped scene-dependent changes in the applied HVS models' magnitude from affecting the IQM scores, but not scene-dependent variations in their shape. It should be noted that IQM variants were also generated without this normalisation, for comparison, to assess whether accounting for the magnitude of scene-dependent HVS models improved their accuracy (the MAE of these variants is analysed in Appendix J).

Output scores from the SQRIn, PIC, log NEQ and Visual log NEQ were calibrated to the SQS₂ scale by:

- 1) setting k_2 to zero, as carried out in a previous comparable investigation [9, p. 60];
- 2) setting the value of k_1 according to Equation 6.2, so that the mean of the IQM scores for all test images at SNR 80 without denoising or sharpening applied, \bar{m}_{output} , equaled the mean of the respective observers' SQS₂ ratings, \bar{m}_{SQS} .

$$k_1 = \frac{\bar{m}_{SQS}}{\bar{m}_{output}} \quad (6.2)$$

The CPIQ metric was calibrated to the SQS₂ scale by subtracting the quality loss in JNDs (i.e. QL_m in Equation 3.7) from the SQS₂ value of 23 corresponding to the input images to the simulation.

6.2 Results

Subjective image quality ratings from the evaluations are presented in Section 6.2.1. IQM variants are benchmarked in terms of their Mean Absolute Error (MAE) and Spearman's Rank Order Correlation Coefficient (SROCC) in Section 6.2.2 and analysed further in Section 6.2.3. Appendix J is also referred to that benchmarks the MAE of every IQM variant.

6.2.1 Subjective Quality Ratings

Figure 6.5 presents uncalibrated subjective quality ratings for each test image. This was because Allen [102, p. 257] observed that calibrating such ratings according to the average scene relationship, as described in ISO 20462 [56, p. 19], removed virtually all scene susceptibility and/or scene-dependency from the data. Thus, uncalibrated SQS₂ ratings were more suitable for evaluating the IQMs of this thesis, which are designed for scene-dependent systems. Uncalibrated ratings were employed successfully in Allen's evaluations of IQM performance, with respect to scene-dependent Joint Photographic Experts Group (JPEG) and JPEG 2000 compression [102, Sec. 7.5].

Figures 6.6 and 6.7 analyse the scenes' susceptibilities and observers' sensitivities to the capture system simulation's artefacts, respectively, following the method of Keelan [46, Ch. 10]. Scenes/observers were grouped in terms of their susceptibility/sensitivity in the following manner. 25% of the scenes/observers with the most data points between the 75% and 100% percentiles were classified as high susceptibility/sensitivity. Likewise, 25% of the scenes/observers with the most data points between the 0% and 25% percentiles were classified as low susceptibility/sensitivity. The remaining 50% of the scenes/observers were classified as medium susceptibility/sensitivity. Since noise and ISP artefacts were more perceptually significant at lower SNRs, the number of data points at SNRs 10, 20, 40 and 80 were weighted by factors of 2.5, 2, 1.5 and 1, respectively, during these classifications.

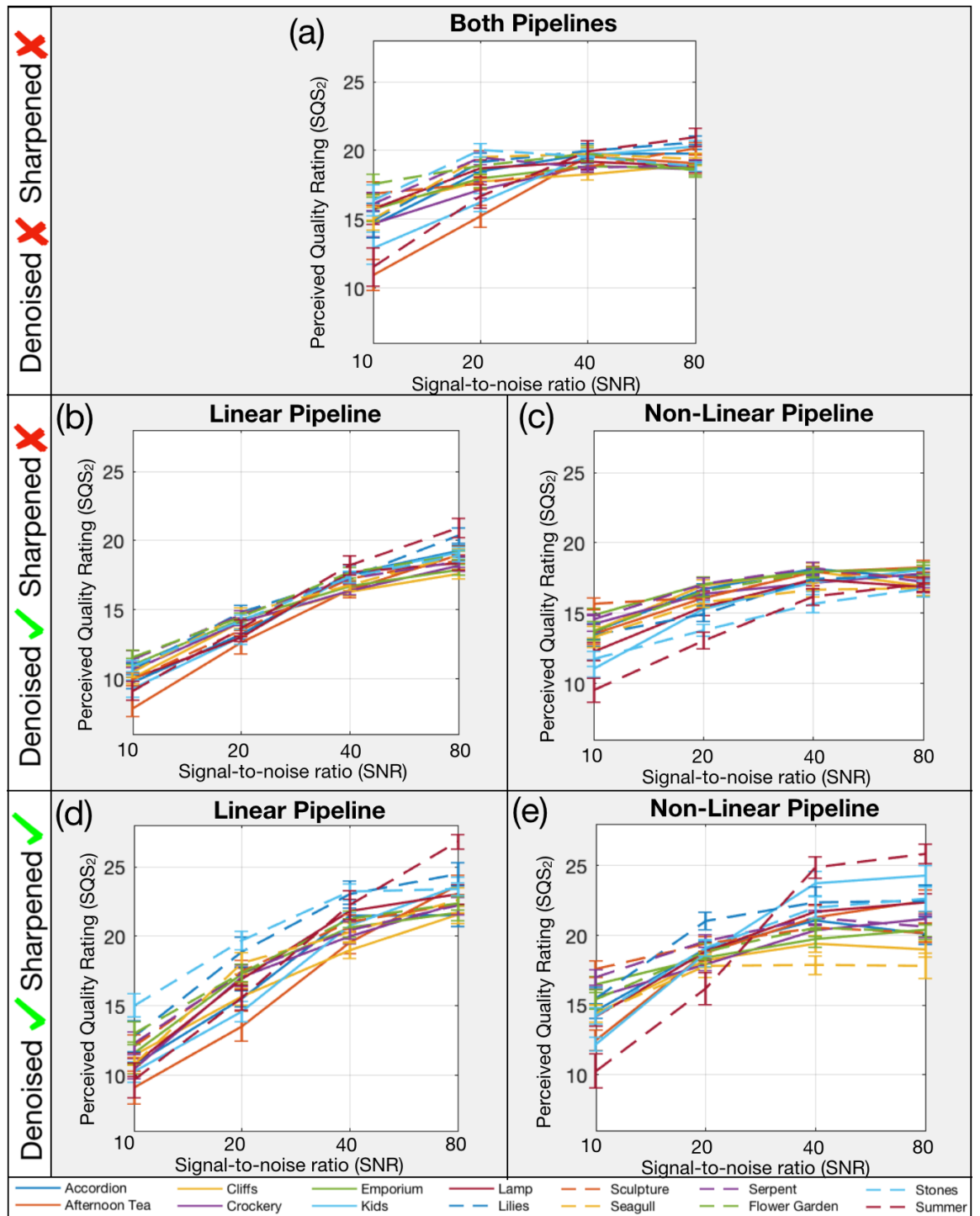


Figure 6.5 Mean observer image quality ratings on the Standard Quality Scale (SQS₂) for each test image; error bars show standard error.

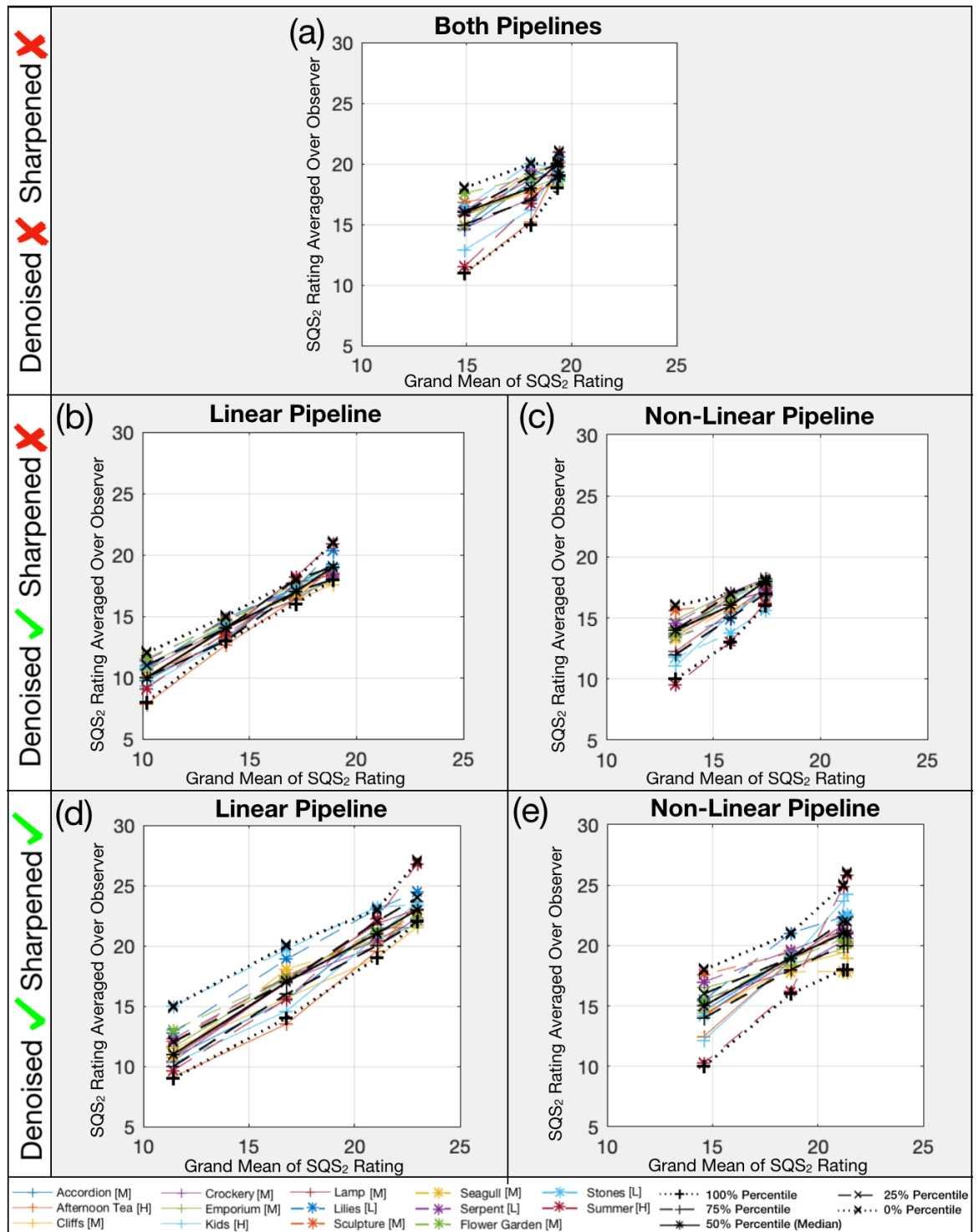


Figure 6.6 The scenes' susceptibilities to perceived quality loss. The x-axis is the grand mean of the scenes' quality ratings. The y-axis is the quality rating for each scene, averaged over all observers. Higher susceptibility scenes are distributed closer to the x-axis. The scenes classified as high, medium and low susceptibility are labelled [H], [M] and [L], respectively.

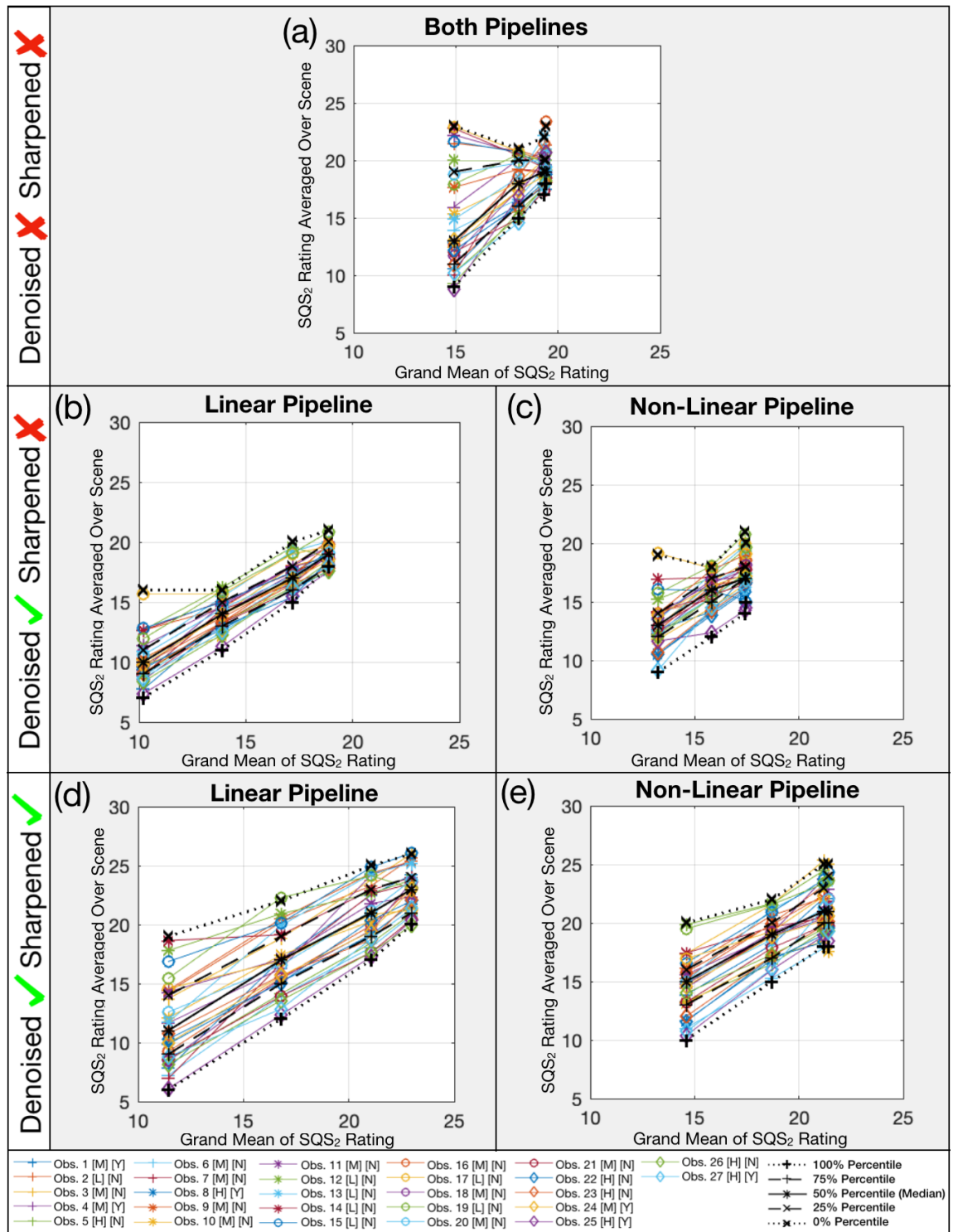


Figure 6.7 The observers' sensitivities to perceived quality loss. The x-axis is the grand mean of the observers' quality ratings. The y-axis is the quality rating for each observer, averaged over all scenes. Data from higher sensitivity observers is distributed closer to the x-axis. Observers classified as high, medium and low sensitivity are labelled [H], [M] and [L], respectively. Experienced and inexperienced observers are labelled [Y] and [N], respectively.

Perceived image quality was increased by combined denoising and sharpening, for which the tuning of the simulations was optimised (see Section 6.1.1 for further information). However, it was reduced by denoising alone. Non-linear denoising and sharpening algorithms preserved image signal content and mitigated the amplification of noise, respectively, as described in Section 2.2.1. Thus, they produced higher quality images than equivalent linear ISP algorithms when activated at a higher intensity at lower SNRs. The perceived quality of images from the non-linear pipeline was slightly more scene-dependent than the linear pipeline.

The variance of the scenes' susceptibilities and observers' sensitivities to quality loss generally increased at lower SNRs. The former was also lower than the latter as observed by Keelan [46, Ch. 10]. Non-busy scene images, e.g. Summer, Afternoon Tea and Kids, were most susceptible to quality losses due to noise and denoising artefacts that dominated at low SNRs. The Seagull and Cliffs images were most susceptible to sharpening artefacts. The latter affected high-contrast edges in particular, dominating certain scenes at higher SNRs.

Observers that were sensitive to noise were generally also sensitive to denoising artefacts. The variance of observer sensitivities was highest at low SNRs before denoising, and after combined sharpening and denoising. This supports Persson's [48, p. 61] observation that the variance of observer sensitivities is higher when the test and ruler image artefacts are dissimilar. Inexperienced observers were less consistent, as noted in comparable studies [201]. This is expected to be either due to fatigue, erratic quality consciousness, or less effective training.

Experienced observers were not omitted from the dataset as suggested by Keelan [46, p. 9], because:

- 1) they represented only 22.2% of all observers;
- 2) previous successful studies using the image quality ruler [48, p. 53], [102, p. 225], [201] included higher proportions of them;
- 3) their sensitivity was not found to differ significantly from inexperienced observers with respect to denoising (also noted by Persson [48, p. 62]) or sharpening artefacts.

6.2.2 Benchmarking of IQM Variants

All IQM variants were benchmarked according to their MAE and root mean square error (RMSE) with respect to the observer image quality ratings, once the IQMs were calibrated. The MAE describes specifically the mean difference in SQS_2 units between the output scores of the metric and the ideal linear relationship with the observer ratings. This ideal linear relationship is indicated by the pink line in Figures 6.10 to 6.16, which has equation $y = x$. Thus, if a metric variant has an MAE of 2, then it can be expected to predict the quality of a given image with an accuracy of ± 2 JNDs. The RMSE is also calculated with respect to the ideal linear relationship. Lower RMSE and MAE values indicate higher IQM accuracy.

Further, all IQM variants, as well as the Mean Structural Similarity (MSSIM) metric, were benchmarked according to their SROCC between the IQM scores and the observer ratings. The SROCC describes the goodness of fit to a monotonic function that may not be linear; higher values indicate a higher correlation, but not necessarily higher accuracy.

The MATLABTM implementation by Wang *et al.* [333] was employed to compute the MSSIM (Equation B9) with respect to the image luminance channel using the default Gaussian window and dynamic range, specified in Appendix B. The constants K_{SSIM_1} and K_{SSIM_2} in Equations B1 and B3 were set to their default values of 0.01 and 0.03, respectively [114]. The reference image was the input image to the simulations. The test image was the output image generated by the pipeline at the required processing stage, at the specified SNR.

Tables 6.4 and 6.5 benchmark the metric variants' MAEs and SROCCs, respectively, with respect to the full dataset of observer quality ratings, i.e. all observers and all scenes. Input parameters for the highest and lowest performing variants are listed in the left-hand and right-hand tables, respectively. RMSE and MAE scores are colour coded across both tables in each figure, from red to green, indicating lowest to highest performance, respectively. Figures 6.8 and 6.9 describe the distribution of the MAE and SROCC scores across all variants of each metric, respectively.

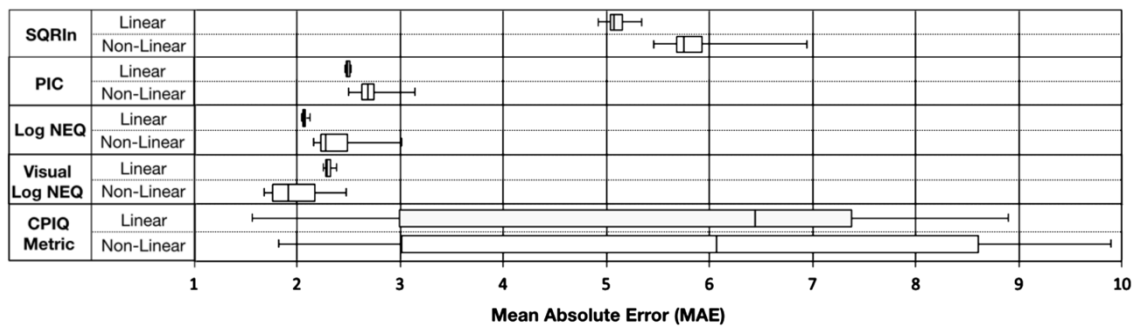


Figure 6.8 Box and whisker plots of Mean Absolute Errors (MAE) of all variants of each metric.

Highest Accuracy Variants of Each Metric						Lowest Accuracy Variants of Each Metric					
Metric Name	Input Parameters			Pipeline Type	MAE	Metric Name	Input Parameters			Pipeline Type	MAE
	Noise Measure	MTF	CSF				Noise Measure	MTF	CSF		
CPIQ	Uses Pictorial Image Replicates	Pictorial Image SPD-MTF	Barten	Non-Linear	1.82	CPIQ	Uses Dead Leaves Replicates	Direct Dead Leaves MTF	cVPF	Non-Linear	9.88
				Linear	1.57					Linear	8.82
Visual Log NEQ	Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	Non-Linear	1.68	Visual Log NEQ	Uniform Patch NPS	Pictorial Image SPD-MTF	cVPF	Non-Linear	2.48
				Linear	2.25					Linear	2.38
Log NEQ	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	N/A	Non-Linear	2.16	Log NEQ	Uniform Patch NPS	Pictorial Image SPD-MTF	N/A	Non-Linear	3.00
				Linear	2.04					Linear	2.13
PIC	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	2.50	PIC	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	3.15
				Linear	2.47					Linear	2.53
SQRIn	Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	5.46	SQRIn	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	6.94
				Linear	4.92					Linear	5.34

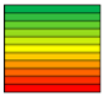

 ← Lowest MAE of All Variants of All IQMs (i.e. Highest Accuracy)
 ← Highest MAE of All Variants of All IQMs (i.e. Lowest Accuracy)

Table 6.4 Input parameters and Mean Absolute Errors (MAE) for the most accurate (left table) and least accurate (right table) variants of each image quality metric (IQM). The MAE scores are coloured from red to green denoting lowest to highest accuracy.

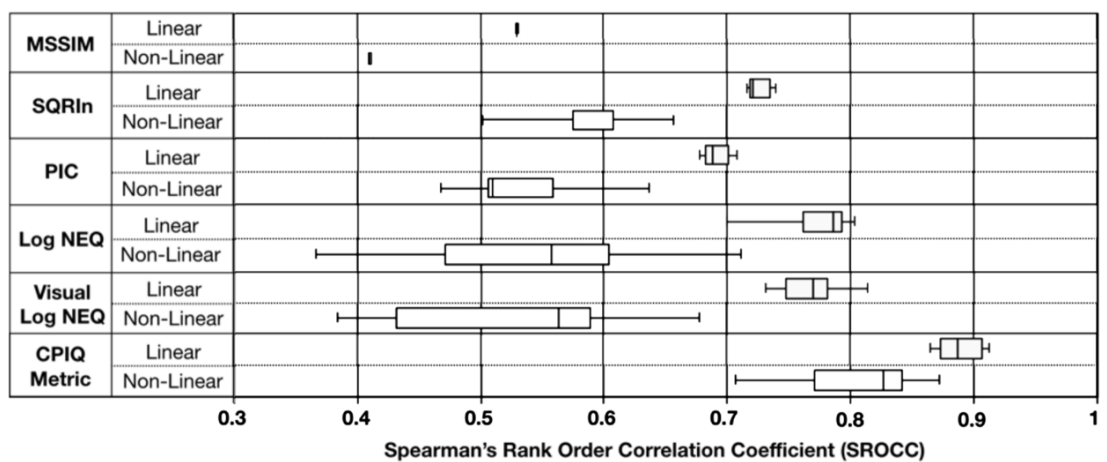



Figure 6.9 Box and whisker plots of the Spearman's Rank Order Correlation Coefficient (SROCC) of all variants of each metric, including the Mean Structural Similarity (MSSIM).

Highest Correlating Variants of Each Metric						Lowest Correlating Variants of Each Metric					
Metric Name	Input Parameters			Pipeline Type	SROCC	Metric Name	Input Parameters			Pipeline Type	SROCC
	Noise Measure	MTF	CSF				Noise Measure	MTF	CSF		
CPIQ	Uses Dead Leaves Replicates	Direct Dead Leaves	Standard	Non-Linear	0.87	CPIQ	Uses Pictorial Image Replicates	Pictorial Image SPD-MTF	Barten	Non-Linear	0.71
				Linear	0.91					Linear	0.87
Visual Log NEQ	Mean Pictorial Image SPD-NPS	Direct Dead Leaves	Barten	Non-Linear	0.68	Visual Log NEQ	Uniform Patch NPS	Mean Pictorial Image SPD-MTF	cVPF	Non-Linear	0.33
				Linear	0.81					Linear	0.73
Log NEQ	Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	N/A	Non-Linear	0.71	Log NEQ	Uniform Patch NPS	Mean Pictorial Image SPD-MTF	N/A	Non-Linear	0.37
				Linear	0.80					Linear	0.70
PIC	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	0.64	PIC	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	0.47
				Linear	0.71					Linear	0.68
SQRIn	Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Any of Barten, cCSF or cVPF	Non-Linear	0.66	SQRIn	Uniform Patch NPS	Direct Dead Leaves MTF	Any of Barten, cCSF or cVPF	Non-Linear	0.50
				Linear	0.74					Linear	0.72
MSSIM	N/A	N/A	N/A	Non-Linear	0.40	MSSIM	N/A	N/A	N/A	Non-Linear	0.40
				Linear	0.53					Linear	0.53



← Highest SROCC of All Variants of All IQMs (i.e. Highest Correlation)

← Lowest SROCC of All Variants of All IQMs (i.e. Lowest Correlation)

Table 6.5 Input parameters and Spearman's Rank Order Correlation Coefficients (SROCC) for variants of each image quality metric (IQM) that resulted in the highest (left table) and lowest correlation (right table), as well as the Mean Structural Similarity (MSSIM) metric. The SROCC scores are coloured from red to green denoting lowest to highest correlation, respectively.

When evaluating the robustness of each metric to changes in its input parameters, the main factors to be considered are the accuracy (MAE) of its most accurate variants, and the range of accuracies across all variants of the metric. The maximum level and range of the variants' correlations with the observer ratings (SROCC), are further factors to be taken into account.

The CPIQ metric produced the variants with highest accuracy overall, followed closely by the Visual log NEQ and log NEQ metrics proposed in this thesis. However, the Visual log NEQ produced the highest accuracy variant for the non-linear pipeline. This is an exciting result, considering it is a very simple metric that relates directly to the fundamental NEQ measure, and accounts for system scene-dependency. The range of accuracies (MAE) of variants of the CPIQ metric was extremely high compared to the other metrics tested, as demonstrated in Figure 6.8. This is expected to be due to the metric's pre-calibration making it sensitive to any changes in its input parameters, particularly the CSF parameter. The simpler Signal Transfer Visual IQMs (STV-IQM) such as the PIC, SQRIn, log NEQ and Visual log NEQ, showed greater consistency when their input parameters were changed to account for imaging system and visual scene-dependency.

The most accurate IQM variants (shown in Table 6.4) all implemented noise measures derived from pictorial images that were most sensitive to system scene-dependency. They also used SPD-MTFs that were sensitive to system scene-dependency, measured either from pictorial or dead leaves signals. They employed the Barten CSF [186] for all IQMs where changing the CSF affected the IQM accuracy significantly. Changing the CSF parameter of the PIC and SQRIn did not affect the accuracy or correlation of the metrics since the high display luminance rendered visual noise very low relative to the imaging system's noise at most SNRs. Thus, the CSF parameter effectively cancelled itself out, since it was applied to both the numerator and denominator of the integral in these IQMs (which are defined in Equations 3.3 and 3.4).

The least accurate IQM variants implemented most often the standard uniform patch NPS and the direct dead leaves MTF [19] that accounted least for system scene-dependency, as well as the cVPF [8] visual model. The pictorial image SPD-MTF was also used by some of the least accurate IQM variants, expected to be due to bias discussed in Chapter 5.

Benchmarking the IQM variants by their SROCC (Table 6.5) showed comparable trends to the MAE benchmarking tables (Table 6.4). The MSSIM provided the lowest SROCC of all. Its RMSE and MAE could not be calculated since it is of range $0 \leq MSSIM \leq 1$.

Appendix J benchmarks the accuracy of every IQM variant, applying the same colour coding as Tables 6.4 and 6.5. MAE scores have been tabulated for various scene and observer groups, as well as for all observers and all scenes under the following conditions:

- 1) without normalising the CSFs following the method of Section 6.1.3;
- 2) after restricting the range of integration to $12 < u \leq \infty$ cycles/degree, to mitigate the effects of bias in the SPD-MTF, discussed in Chapter 5.

Results from these various data subsets were generally in agreement with one another. Trends in the MAE tables were virtually identical to the equivalent RMSE tables and shared commonalities with equivalent SROCC tables, which were both omitted from this thesis.

Benchmarking tables in Appendix J, along with Tables 6.4 and 6.5 and observations from Sections 4.3 and 5.2, indicate that the various SPD-MTF and SPD-NPS measures accounted well for system scene-dependency. The IQM accuracy increased consistently for the non-linear pipeline when these measures were used and was sometimes even higher than for the linear pipeline. However, the IQMs that implemented SPD-MTFs from pictorial images

were, in some cases, less accurate than IQMs that used dead leaves SPD-MTFs or the direct dead leaves MTF. This observation is expected to be due to bias in the former measurements. Restricting the range of integration to $12 < u \leq \infty$ cycles/degree did not improve IQM accuracy, despite excluding frequencies with greatest bias and preserving the fundamental spatial frequencies for seeing objects [334].

6.2.3 Further Analysis of Selected IQM Variants

This section analyses the correlations between selected variants of metrics from this thesis and the observer image quality ratings. This analysis is intended to complement the benchmarking tables of Section 6.2.2 and Appendix J.

Figures 6.10 to 6.16 illustrate one or more of the following, concerning each IQM of this thesis:

- 1) typical IQM variant behaviour;
- 2) best/worst case scenarios regarding IQM variant accuracy;
- 3) significant changes in IQM variant accuracy that resulted from using input parameters that were more sensitive to scene-dependency.

The data presented is from the full dataset of observers and scenes. Figure 6.17 analyses the MSSIM's correlation. Data in Figures 6.11, 6.12, 6.15 and 6.16 is coloured with respect to each scene to express scene-dependency in IQM behaviour. To isolate the effects of different ISP stages in Figures 6.10, 6.13, 6.14 and 6.17, data points generated before and after denoising are coloured red and green, respectively, and data generated after denoising and sharpening is coloured blue; darker markers indicate higher SNRs.

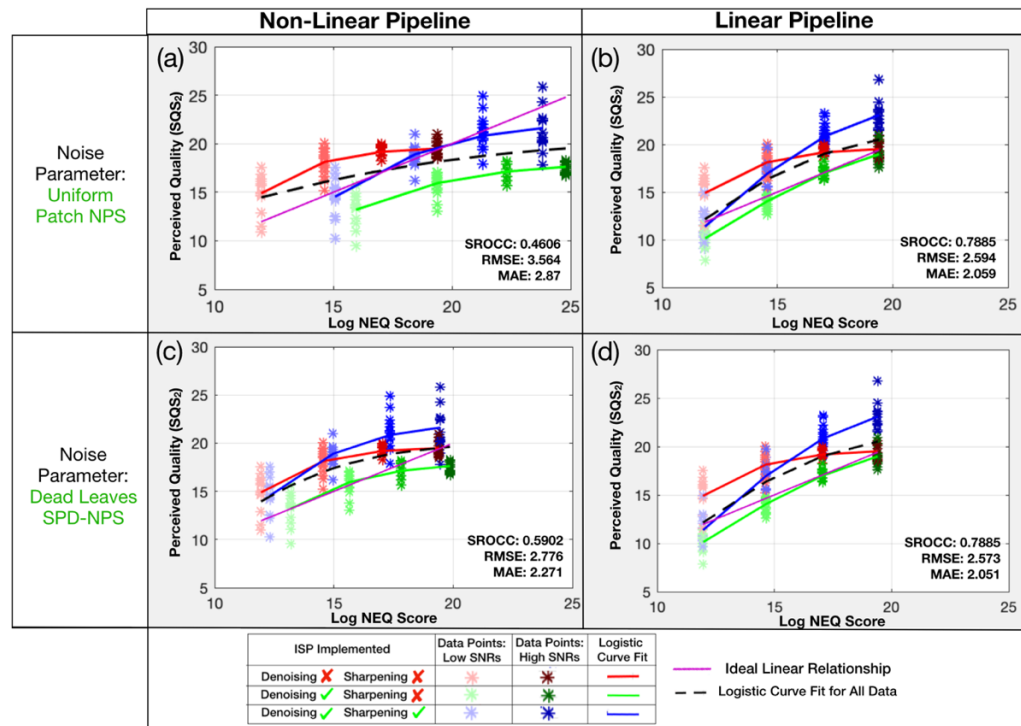


Figure 6.10 The mean of all observer image quality ratings vs the log Noise Equivalent Quanta (NEQ) metric variants that were less accurate or sensitive to system scene-dependency. They implement direct dead leaves Modulation Transfer Functions (MTF) [19].

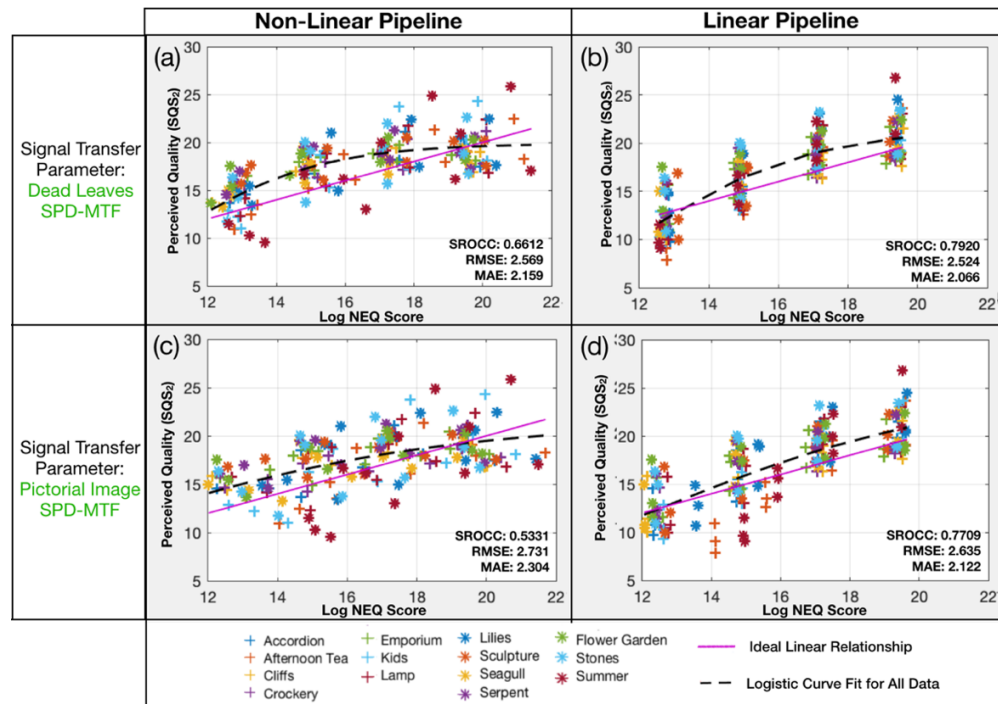


Figure 6.11 The mean of all observer image quality ratings vs log Noise Equivalent Quanta (log NEQ) variants that were more accurate and sensitive to system scene-dependency. They employ the pictorial image SPD-NPS. The most accurate log NEQ variant is shown in (a) and (b).

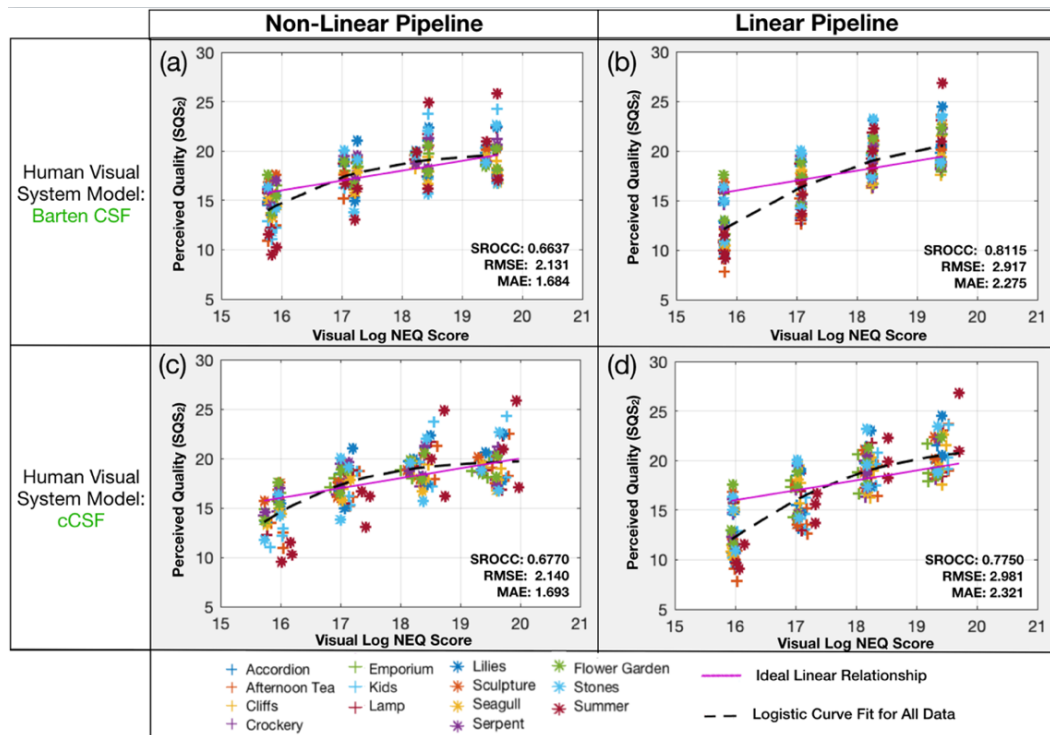


Figure 6.12 The mean of all observer image quality ratings vs the most accurate Visual log NEQ variants that both implemented the dead leaves SPD-MTF and mean pictorial image SPD-NPS.

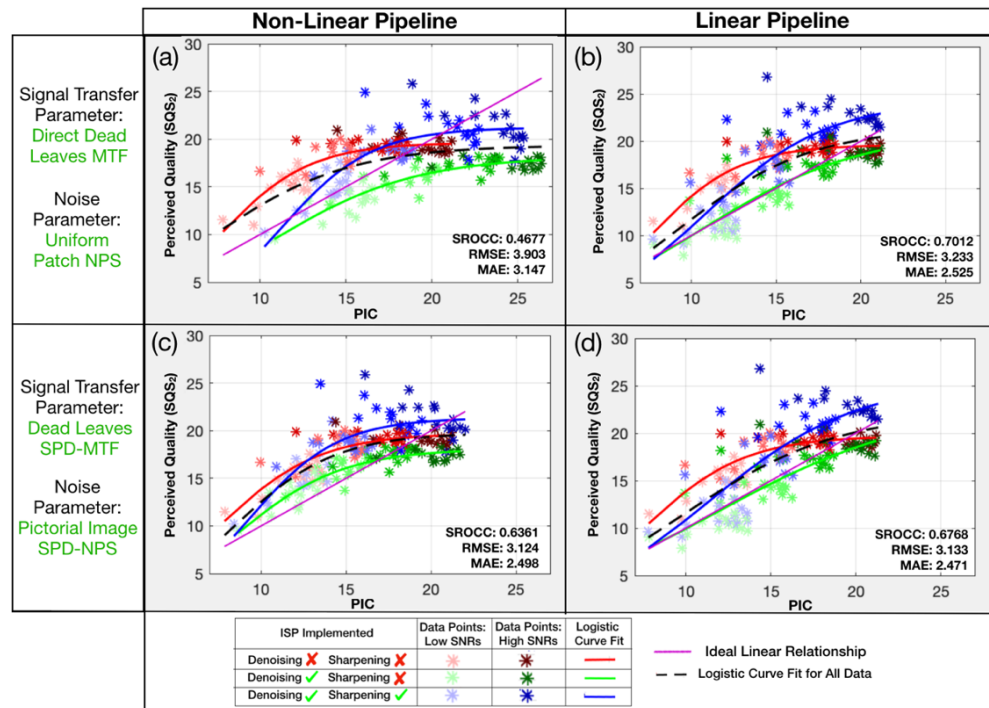


Figure 6.13 The mean of all observer quality ratings vs the Perceived Information Capacity (PIC) [169] variants with lowest, (a) and (b), and highest accuracy, (c) and (d). Implementing different visual models had minimal effect on the correlations; hence the employed model is not specified.

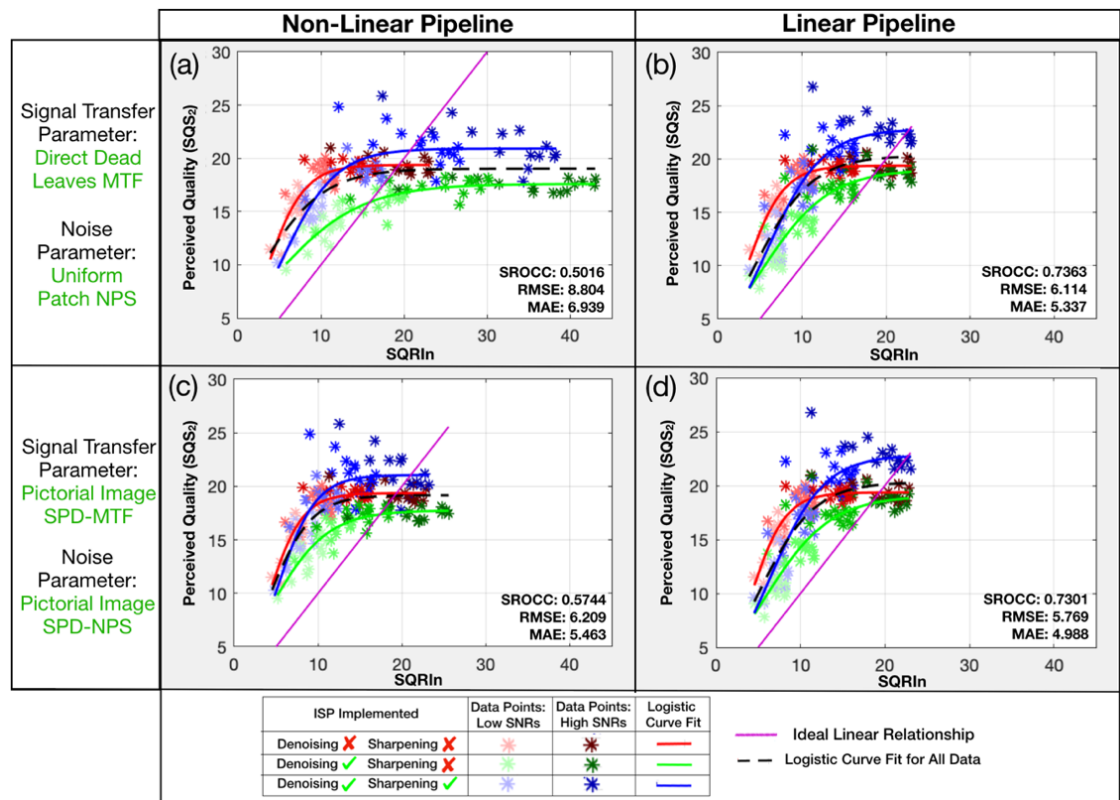


Figure 6.14 The mean of all observer image quality ratings vs the Square Root Integral with noise (SQRIn) [12] variants of lowest, (a) and (b), and highest accuracy, (c) and (d). The visual model is not specified because implementing different models had minimal effect on the correlations.

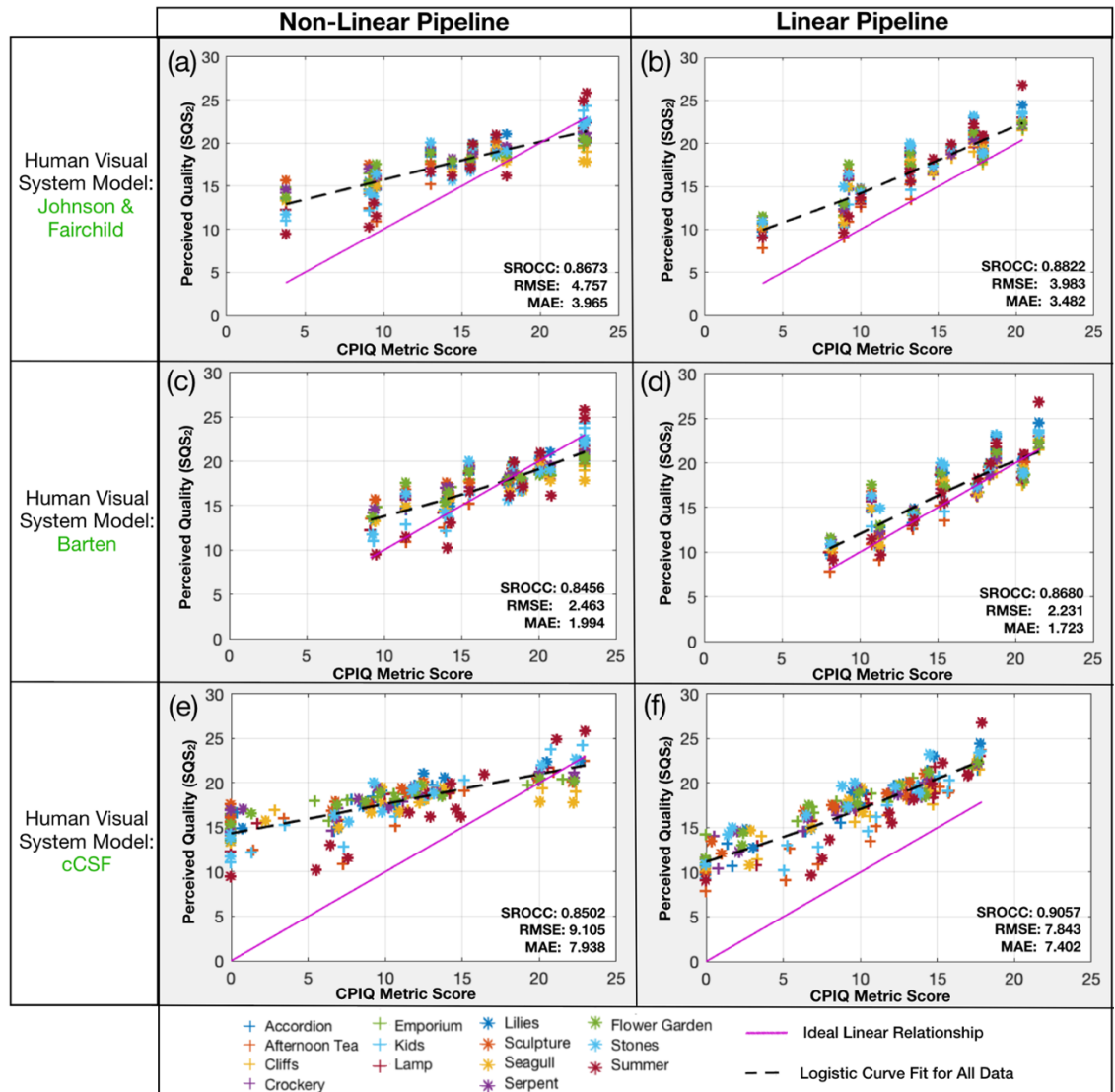


Figure 6.15 The mean of all observer quality ratings vs CPIQ metric [22] variants that employed different contrast sensitivity functions (CSF). All variants implemented the direct dead leaves Modulation Transfer Function (MTF) [19] and uniform patch noise images that are default input parameters to the IEEE P1858 standard [22] and are least sensitive to system scene-dependency.

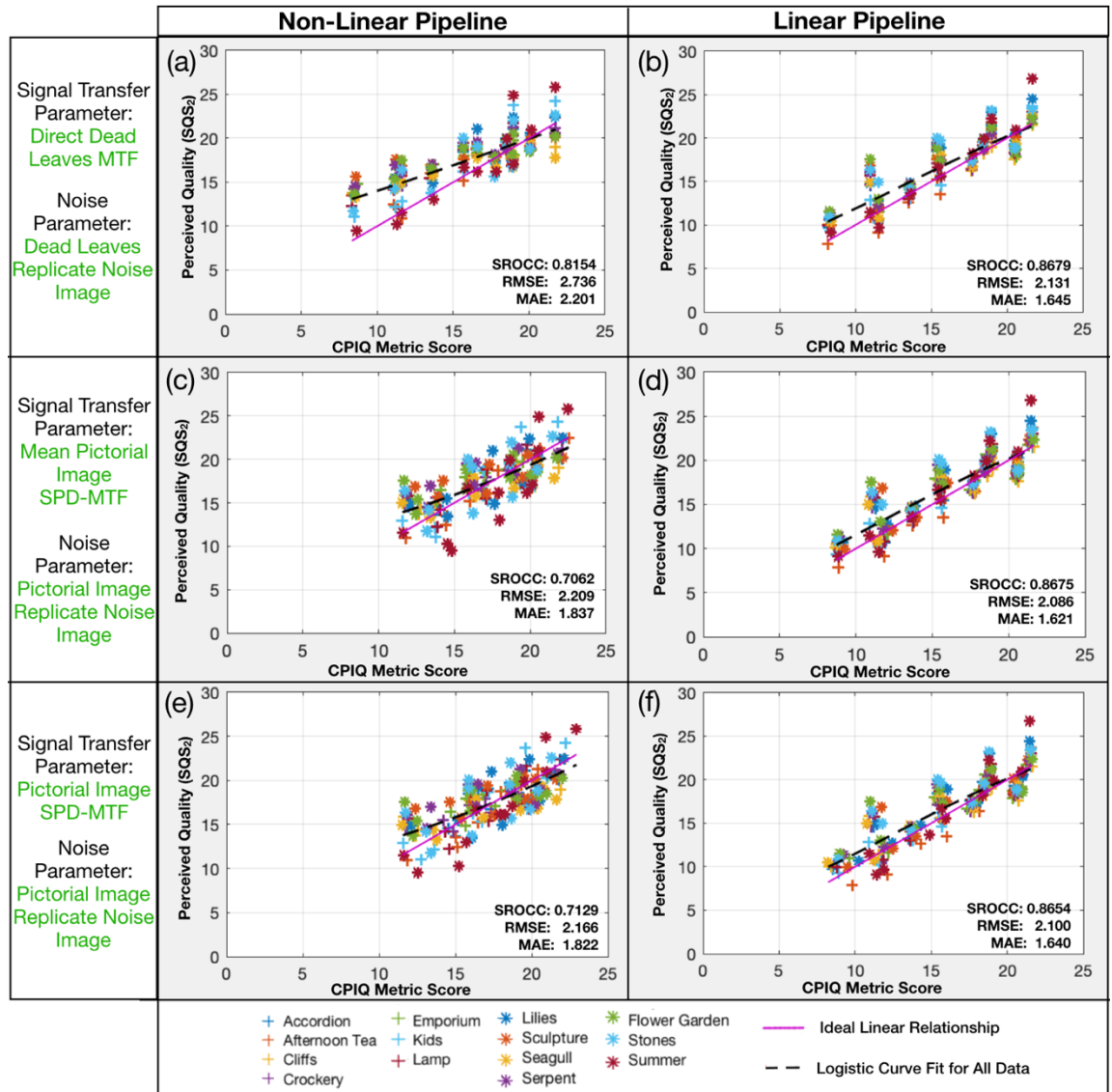


Figure 6.16 The mean of all observer image quality ratings vs CPIQ metric [22] variants that employed different Modulation Transfer Functions (MTF) and noise parameters. All variants implemented the Barten contrast sensitivity function (CSF) [186] that was most optimal for the CPIQ metric. The most accurate CPIQ metric variant is shown in (e) and (f).

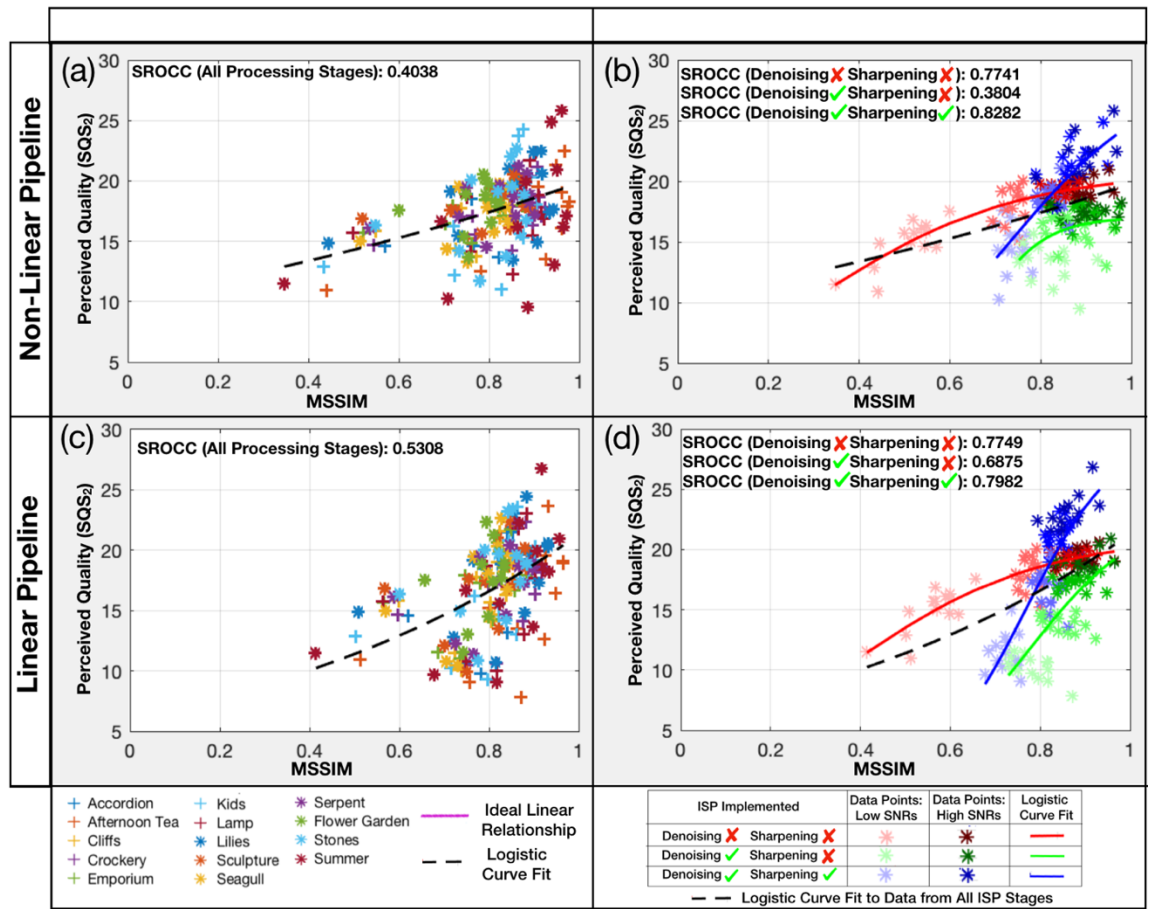


Figure 6.17 The mean of all observer quality ratings vs output scores from the Mean Structural Similarity (MSSIM) metric [114].

Variants of the PIC [169] and SQRIn [12] displayed comparable trends to variants of the log NEQ and Visual log NEQ in Figures 6.10 to 6.14, as well as Tables 6.4 and 6.5 and Appendix J. This is expected to be because the former STV-IQMs are fundamentally signal-to-noise measures and are thus also closely related to the NEQ (as noted in Section 2.4).

But the CPIQ metric [22] displayed contrasting behaviour which was expected to be due to:

- 1) differences in design, since it is a more complex Multivariate Formalism IQM (MF-IQM; see Section 3.1.2 for more information);
- 2) it relating less closely to the NEQ;
- 3) it applying more extensive pre-calibration (as expressed in Figure 3.3).

The effect of implementing the revised noise, signal transfer and CSF input parameters in the CPIQ metric are expected to be distorted because the calibration was tuned for the original parameters of the IEEE P1858 CPIQ standard [22], including uniform patch noise images, the direct dead leaves MTF [19], and Johnson and Fairchild's CSF [188].

The following observations demonstrate why the accuracy of the PIC, SQRI_n, log NEQ and Visual log NEQ improved significantly for the non-linear pipeline when SPD-NPS measures were employed that accounted for system scene-dependency (as demonstrated in Tables J2 to J6). Since Figure 4.5 demonstrates that the uniform patch NPS underestimated noise after non-linear denoising, the IQMs that employed this measure overestimated the observer SQS₂ ratings (Figures 6.10(a), 6.13(a) and 6.14(a)). This overestimation of perceived quality was mitigated by implementing the dead leaves SPD-NPS (Figure 6.10(c)) and to a greater extent the pictorial image SPD-NPS (Figures 6.13(c) and 6.14(c)) that characterised more appropriately the non-linear pipeline's noise.

The SQRI_n overestimated the observed SQS₂ ratings at high SNRs, forming a curved distribution with decreasing gradient. This supports Töpfer and Jacobson's [169] observation that the SQRI_n does not always change linearly with noise or describe perceived image quality in equal JND units, as discussed in Section 3.1.1. The expected cause of this limitation is that it underestimates the perceptual impact of noise at near-threshold levels [169]. The PIC also displayed such behaviour, but to a lesser extent. If the SQRI_n or PIC were to be re-calibrated at SNR 20 and data at SNRs 40 and 80 were omitted, Figures 6.13 and 6.14 suggest they would be more representative of the ideal linear relationship and their accuracy would improve significantly. Thus, it is concluded that the simulations at SNRs 80 and 40 reached the limit to which the SQRI_n and PIC apply, with respect to noise.

The CPIQ metric's distribution was linear when plotted versus the observer SQS₂ ratings and was close to the ideal linear relationship for the most accurate variant, which employed SPD-MTFs and noise images derived from images of pictorial scenes (Figure 6.16(e) and (f)). Changing the CSF function altered the gain and offset of this linear distribution, affecting the accuracy of the IQM significantly. The gain and offset was most optimal when the Barten CSF [186] was employed. This function has a similar band-pass shape to the Johnson and Fairchild CSF [188] from the IEEE P1858 CPIQ standard [22] but displayed minor scene-dependent variations in shape as indicated by Figure I5. Implementing the cCSF [35] or cVPF [8] produced unfavourable gain and offset. These two highly scene-dependent HVS models were generally slightly flatter in profile and, once normalised, prioritised high frequencies more than the CSF from the standard [188]. This likely rendered them less compatible with the calibration or curve fitting of the CPIQ metric (as discussed further in Section 7.4). Thus, their negative impact on the accuracy of this metric may not be indicative of their appropriateness as HVS functions for image quality modelling.

The CPIQ metric's accuracy improved, however, when SPD-MTFs and noise images from pictorial images were employed (Figure 6.16), despite the limitations of revising the input parameters of IQMs that applied pre-calibration or curve fitting (discussed further in Section 7.1).

This observation, alongside their successful implementation in other metrics, demonstrates the following for the various SPD-MTF and SPD-NPS measures:

- 1) They are robust.
- 2) They are more relevant to image quality modelling than current standard system performance measures.
- 3) They can likely be substituted successfully in the place of current standard measures, even within IQMs implementing pre-calibration or curve fitting.

The accuracy of the MSSIM was limited by the compound effect of blur, noise, demosaicing, denoising and sharpening artefacts. Further work should investigate the extent to which these limitations apply to other computational IQMs (CP-IQM) (Section 8.2).

6.3 Summary

Correlations between observer image quality ratings (measured using the ISO 20462 image quality ruler) and a number of variants of several IQMs have been evaluated. Variants of each IQM used different combinations of input parameters. They accounted for scene-dependency of the imaging and human visual system to varying degrees. The conclusions of this chapter are core to the discussions of the thesis. They support the hypothesis that imaging system scene-dependency should be accounted for by IQMs, and further validate the various SPD-NPS and SPD-MTF measures from pictorial image and dead leaves signals. Less support was found for accounting for spatial visual scene-dependency, although this observation should not be considered conclusive and should be investigated further. The novel log NEQ and Visual log NEQ metrics that are a product of this research, as well as the revised scene-and-process-dependent CPIQ metric, performed particularly well when they implemented the SPD-MTF and SPD-NPS measures. The latter metric was less consistent.

Chapter 7 Discussion

Spatial image quality metrics (IQM) were firstly reviewed from an image capture systems development viewpoint in this thesis. Metrics from the Signal Transfer Visual IQM (STV-IQM) and Multivariate Formalism (MF-IQM) genres were concluded to be most suitable for the purpose. These metrics use as input parameters the Modulation Transfer Function (MTF) and Noise Power Spectrum (NPS), which are standard measures of signal transfer and noise, respectively; both are used routinely when developing capture systems. The most suitable metrics also implemented contrast sensitivity functions (CSF) describing human visual sensitivity under the given viewing conditions. Prior art demonstrated these metrics are less accurate for the characterisation of non-linear systems. The hypothesis in this thesis is that, these inaccuracies are due to limitations of current MTFs, NPSs and CSFs, presented below.

- 1) Firstly, non-linear capture system scene-dependency is unaccounted for in MTF and NPS measurements derived from test targets that contain uniform tone patches, edges, sine-waves and white noise. Also, since none of these signals are representative of the “average scene”, the derived measurements do not describe the average real-world performance of such systems.
- 2) Secondly, the traditional CSF (which is a model of detection of unmasked narrow-band signals) is not an accurate model for describing the visual tasks related to image quality evaluation. The latter is expected to involve both detection and discrimination of complex image signals that are usually masked; both were found to be scene-dependent in prior art [8], [35].

Frameworks were proposed to revise metrics to incorporate the following MTFs, NPSs and CSFs, with the aim of improving metric accuracy:

- 1) Scene-and-process-dependent NPSs (SPD-NPS) measured from replicate captures of images of scenes (or suitable test charts) to account for noise scene-dependency.
- 2) Scene-and-process-dependent MTF (SPD-MTF) measures derived from images of scenes (or appropriate test charts) that account for signal transfer scene-dependency.

- 3) Contextual CSFs (cCSF) [35] or Contextual Visual Perception Functions (cVPF) [8] that account for the scene-dependent effect of visual masking during detection or discrimination of pictorial scene signals, respectively.

Frameworks for two novel metrics (the log Noise Equivalent Quanta (log NEQ) and Visual log NEQ) were also proposed that used the abovementioned parameters.

The SPD-MTF and SPD-NPS measures were developed in this thesis and validated using camera simulation pipelines that generated controlled outputs. The pipelines applied either linear or non-linear content-aware ISP algorithms, such as demosaicing, denoising and sharpening. The metrics were validated using the same pipelines.

Although the measures and metrics were not validated using real capture systems, the results from these simulations are likely to be representative of real systems. This is particularly the case for observations regarding system scene-dependency, and the capability of the measures and IQMs to account for it. This is because scene-dependent system behaviour related far less to the initial modelling of image capture (i.e. lens blur, noise, mosaicing and pre-processing) compared to the ISP algorithms applied thereafter.

The discussion comprises of five sections. The first four concern the contributions of the four research chapters (Chapters 3 to 6). Section 7.1 discusses the scene-and-process-dependent IQM frameworks. Sections 7.2, 7.3 and 7.4 discuss the SPD-NPS measures, SPD-MTF measures, and scene-and-process-dependent IQMs, respectively. Finally, Section 7.5 reflects upon the implementation and application of all measures and metrics of this thesis.

7.1 Image Quality Metric Frameworks

The IQM frameworks were proposed in Chapter 3. The concept behind them was to transfer to STV-IQMs and MF-IQMs the capability of Computational IQMs (CP-IQM) to account for imaging system and human visual scene-dependency. To achieve this, the “nature” of the input parameters of various STV-IQMs and MF-IQMs was revised to analyse contents from the output image (i.e. a captured scene). Namely, their MTF, NPS and CSF parameters were substituted for SPD-MTFs, SPD-NPSs and cCSFs/cVPFs, respectively. The novel log NEQ and Visual log NEQ metric frameworks were developed following the same principle.

The resultant IQM frameworks represent a new sub-genre of STV-IQMs and MF-IQMs. They are adaptive to both imaging and visual system scene-dependency. But they still remain modular and relate to the underlying physics of both systems, and the viewing conditions. The metric frameworks demonstrate some novelty on a conceptual level.

Substituting the input parameters of existing STV-IQMs and MF-IQMs did not violate these metrics, provided the metrics were based upon fundamental relationships between perceived image quality and the MTF, NPS, Noise Equivalent Quanta (NEQ) and/or CSF. However, this does not apply to metrics that involve significant pre-calibration, or constants that have been based on visual image quality rating datasets. Revising the input parameters of such metrics can render their original calibration (or curve fitting) less appropriate since it was tuned for their original input parameters. In such cases, this places the revised IQM at a relative disadvantage to its original incarnation. It can also cause the accuracy of the revised IQM to be unrepresentative of the appropriateness of its input parameters. Also, any pre-calibration (or curve fitting) applied to a given IQM is, strictly speaking, only guaranteed to be applicable to the corresponding subjective image quality rating dataset. In certain cases, such metrics can lose accuracy when applied to new applications, or to systems that significantly exceed or fall below the quality range of the original dataset. These two issues were noted for the Square Root Integral with Noise (SQRI_n) [12] and the Perceived Information Capacity (PIC) [169] metrics in Section 6.2.3. Conversely, other investigations found the IEEE P1858 Camera Phone Image Quality (CPIQ) metric could be applied accurately outside the quality range of its original dataset [335].

Strictly speaking, the proposed metric frameworks require a functional prototype of the imaging system in order to be executed. This is because the SPD-MTF and SPD-NPS are computed from captured (or processed) images. Nevertheless, an envelope of SPD-MTFs and SPD-NPSs can be simulated for a given set of input scene images. Likewise, the cCSF and cVPF can be computed with respect to these scene images. The resultant measures can be used in the metrics to predict the quality ranges of systems under development.

7.2 Scene-and-Process-Dependent Noise Power Spectra (SPD-NPS)

The novel SPD-NPS measures proposed in this thesis are deemed to be robust. They are the most relevant measures available for temporally varying noise in systems that apply non-linear content-aware ISP. These observations are supported by the validations of each SPD-NPS measure (Chapter 4), and the fact that the accuracy of IQMs improved when these measures were used (Chapter 6 and Appendix J).

The measures were validated using image capture pipelines that introduced Poisson noise at signal-to-noise ratios (SNR), ranging from very good to very poor exposure conditions. Read noise and dark current noise were simulated as Gaussian noise with increased mean and standard deviation at lower SNRs. This noise modelling accounted for the reduced quantum efficiency of the red and blue pixels, but not fixed pattern noise (FPN).

The validations indicated that the uniform patch NPS is unrepresentative of temporally varying noise in non-linear image capture systems. This was found to affect the accuracy of any IQMs that used it. Aside from its convenience and ability to account for FPN, there is little to support the use of the uniform patch NPS when characterising non-linear systems.

The dead leaves SPD-NPS proposed in this thesis was found to be more appropriate for the image quality modelling of non-linear systems, as predicted. This supports Artmann's [18] observations on the benefits of characterising capture system noise using dead leaves signals. Artmann's [18] noise measures are derived indirectly by comparing measurements from different dead leaves MTF measurement implementations. The dead leaves SPD-NPS has certain limitations (discussed below) but measures the NPS directly, involving fewer assumptions. Artmann's noise measures should be compared with the dead leaves SPD-NPS in further work.

Temporally varying noise in dead leaves signals were, however, processed differently by non-linear content-aware ISP algorithms compared to noise in individual images of scenes, or in the "average scene". The dead leaves chart shares little similarity with pictorial scene images, both objectively and visually, beyond replicating the average scene power spectrum. Its mathematically-generated signal consists of randomly distributed, overlaid discs with "perfect" edges. The target content does not represent the variety of edge gradients, complex

structures and surface textures of pictorial scene images, or the fact that these contents are usually distributed in a non-random, structured fashion. Work in this thesis demonstrated that, the processing of noise by non-linear content-aware ISP algorithms varies significantly between different pictorial scene images, due to variations in their signal content. Thus, it is understandable that these algorithms should process noise in dead leaves signals differently to noise in the “average scene”. This leads to an important question for investigation in further work: *can a single test chart be developed that triggers non-linear content-aware ISP algorithms like the average pictorial scene, and if so, to what extent must it replicate real scene signals?*

The results from Chapters 4 to 6 support, in particular, the relevancy of the SPD-NPS measures that were derived from images of scenes. These measures displayed comparable levels of bias to the uniform patch NPS. They also demonstrated significant advantages over other measures when applied to the non-linear pipeline. For example, scene-dependent variations in the shape of the pictorial image SPD-NPS resulted from it accounting for the effect of the input scene image on the processing of noise by non-linear ISP algorithms. The IQMs that used this measure predicted subjective image quality consistently with the highest accuracy (as demonstrated in Tables J1-J6). This important observation indicates that scene-dependent variation in the temporally varying noise of non-linear systems is both measurable, and relevant to image quality. The apparent robustness of the pictorial image SPD-NPS forms a solid foundation for the pictorial image SPD-NPS standard deviation measure of system scene-dependency. It also supports the case for the mean pictorial image SPD-NPS being the best measure of average real-world noise for non-linear systems. Indeed, the IQMs that used the latter measure were slightly more accurate than metrics using the dead leaves SPD-NPS. This difference in accuracy did not reflect the significant disparity between these two measures’ curve shapes. The reason for this was the different curve shapes did not result in significant variations in integrated area.

The limitations of these new SPD-NPS measures are summarised in Section 4.1, and include:

- 1) the requirement for many replicates to be captured (noise is underestimated if fewer replicates are used);
- 2) their higher computational complexity than existing measures;
- 3) their accuracy being dependent on all replicates being registered accurately;
- 4) their inability to account for FPN or fixed patterns of artefacts.

1) and 2) are becoming increasingly less problematic. Many cameras can now be controlled by software to capture images automatically under pre-specified settings and output them directly to a computer. This feature is common for camera phone and autonomous vehicle camera modules, for which the SPD-NPS measures are deemed to be most useful. Computing the mean of the replicates, and registering them, are the only processes that increase in computational demand with greater numbers of replicates. Thus, SPD-NPSs can be measured using hundreds of replicates to minimise underestimation of noise, which corrupts the SPD-MTF measures and IQMs of this thesis.

Most modern capture systems are mirrorless. This mitigates vibration and misregistration of the replicates (3)), which can also be limited by using an automated robotic arm [336], [337] to optimise and maintain steady alignment of the capture device. Various algorithms [338]–[340] can also correct misregistration automatically in terms of horizontal and vertical translation, scale, and orientation. The fact that these algorithms can distort the original camera signal and noise should not be overlooked, although prior art [7], [161] applied simple x-y translation algorithms successfully to register captured replicates in comparable measurements of noise.

FPN and other forms of capture system sensor noise have been minimised over the last 20 years [167]. FPN is now less significant than temporally varying noise under most conditions. It can be measured from replicate captures of a uniform patch following ISO 15739 [168, p. 15] procedures. The measured FPN power spectrum can be added to the SPD-NPS. The resultant total system noise measure still has the limitation of non-linear ISP processing FPN differently in captured uniform patches compared to images of scenes. But this is minor in comparison to taking all noise measurements from uniform patches.

7.3 Scene-and-Process-Dependent Modulation Transfer Functions (SPD-MTF)

The results from Chapters 5 and 6 demonstrate the potential of the SPD-MTF measures and support their use as IQM input parameters. However, this support was less conclusive than for the SPD-NPS measures for two reasons. The first is the bias in the SPD-MTF measurements due to signal-to-noise limitations (discussed in Chapter 5). The SPD-MTF measures inherited these limitations from the direct dead leaves MTF method; the limitations

affected, in particular, SPD-MTFs measured from pictorial images. The second is that the SPD-MTFs were validated by comparison with the direct dead leaves MTF. The latter presented a more appropriate input signal than the uniform patch NPS against which the SPD-NPSs were validated. It is expected that the SPD-MTFs would have demonstrated greater improvements over MTFs measured from edges, sine-waves or white-noise signals. The same trends would also be expected in the accuracy of IQMs that use these measures.

All IQMs performed more accurately using the dead leaves SPD-MTF than the direct dead leaves MTF (Tables J1-J6). The only exception was the IEEE P1858 Camera Phone Image Quality (CPIQ) metric which has inherent constants that originate from its curve fitting (making it less adaptable). The improvements were minor, but they further demonstrate the relevance of the dead leaves SPD-NPS measure, which is an inherent part of the dead leaves SPD-MTF.

The CPIQ metric and SQRIn were most accurate when they used SPD-MTFs measured from images of scenes. But the accuracy of other metrics was less competitive. Thus, the benefits of the pictorial image SPD-MTF and mean pictorial image SPD-MTF accounting most thoroughly for system scene-dependency were traded off by their bias. Reducing the integration range to $0 \leq u \leq 12$ cycles/degree did not improve metric accuracy (as shown in Tables J1-J6) despite omitting the most biased frequencies. It affected the metrics' capabilities to account for sharpness changes at higher quality levels. But the metrics still accounted for the most relevant frequencies for seeing objects [334].

Despite Chapter 5 indicating that the signal transfer of non-linear ISP algorithms differs for the dead leaves test chart and the average scene, implementing SPD-MTFs measured from images of scenes in IQMs is not presently justified. However, the successful implementation in IQMs of SPD-NPS measures derived from images of scenes, suggests the respective SPD-MTF measures should also improve metric accuracy, provided their bias is mitigated further.

The pictorial image SPD-MTF and mean pictorial image SPD-MTF account for signal transfer concerning various signals, including edges and textures. If their bias were mitigated further, implementing either measure in the P1858 CPIQ standard's [22] texture loss attribute metric should also perform the role of its Edge Spatial Frequency Response (SFR) metric.

A parallel PhD study at the Computational Vision and Imaging Technology group, University of Westminster, is currently deriving SFRs from edges extracted directly from captured scenes [341]. The main difference between these SFRs and the SPD-MTFs of this thesis is the former requires no knowledge of the input signal and does not determine a global MTF/SFR measurement. There has been no comparison between the error, speed and computational complexity of these two very different approaches yet.

7.4 Scene-and-Process-Dependent Image Quality Metrics

The revised and novel IQM frameworks were successful. Many variants of each metric were generated, using different combinations of MTF, NPS and CSF parameters in Tables 4.1, 5.1 and 6.1, respectively. All metric variants were validated with respect to images from image capture pipelines with ISP filters tuned at reduced opacity. This placed the variants that used SPD-MTF and SPD-NPS measures at a disadvantage to the variants that used the uniform patch NPS and direct dead leaves MTF. This tested the robustness of the former variants for the reasons given below.

Bias in the SPD-NPS measures related to the level of noise and the number of replicates. Reducing the opacity of denoising increased the proportion of higher-power, unfiltered noise in the denoised image, biasing all SPD-NPS measures. It increased the accuracy of the uniform patch NPS, however, which measured unfiltered noise more accurately than filtered noise. Reducing the ISP opacity also gave the direct dead leaves MTF an advantage over the various SPD-MTFs. This was because the resultant bias in the SPD-NPS measures was carried into the SPD-MTFs due to the latter's signal-to-noise limitations (Equation 5.1).

Despite this handicap, metric variants that used SPD-MTFs and SPD-NPSs were consistently more accurate than variants using the direct dead leaves MTF and uniform patch NPS. This observation demonstrates the robustness of the SPD-MTF and SPD-NPS measures, and their relevance to the characterisation and image quality modelling of non-linear capture systems. Further, it can be inferred that if the ISP filters were at full opacity, the metric variants that used SPD-MTF and SPD-NPS measures would have outperformed the others by a greater margin.

The STV-IQMs complement metrics from the more complex MF-IQM genre, such as the CPIQ metric. The novel log NEQ and Visual log NEQ metrics developed in this thesis are

STV-IQMs that are expected to be valuable to the field. Their comparatively high accuracy and consistency supports Keelan's [176] statement on the fundamental relevance of the NEQ to imaging systems characterisation and image quality modelling. It also supports the range of prior art that applies the NEQ in models of the ideal observer with respect to signal fidelity in the presence of noise [179], [180], and for signal detection [342] under various capture conditions. The fact that these metrics involved minimal calibration or curve fitting meant that analysing their accuracy informed usefully on the relevance of their input parameters to image quality modelling. These metrics consistently performed most accurately using the various SPD-MTF and SPD-NPS measures. This observation was particularly the case for the non-linear pipelines (for which these measures and metrics were designed).

It is expected that applying further calibration or curve fitting would have improved the accuracy of the log NEQ and Visual log NEQ with respect to the subjective image quality dataset of this thesis. However, it is likely that the metrics would then:

- 1) relate less directly to the fundamental NEQ measure and Fechner's law;
- 2) inform less accurately regarding the relevance of their input parameters to image quality modelling;
- 3) be less broadly applicable or revisable;
- 4) potentially be less accurate outside of the range of the dataset to which their calibration or curve fitting was applied;
- 5) inherit certain characteristics of the systems they were calibrated with respect to.

Thus, it was decided that minimal calibration and no curve fitting should be applied, to retain the metrics' "purity" and maximise their relevance across different systems and applications.

The SQRIn and PIC were not as accurate as the "purer", simpler, log NEQ and Visual log NEQ metrics. This is expected to be because they were calibrated with respect to datasets generated by analog capture systems of lower quality. It is proposed that they could be re-calibrated to represent the performance of higher-quality digital capture systems better.

The highest performing variants of the CPIQ metric used SPD-MTF measures and scene-and-process-dependent noise images. They outperformed all variants of the other IQMs despite the CPIQ metric's curve fitting hindering any revisions of it. However, the following unexpected behaviour was also noted. Using scene-and-process-dependent noise images in the CPIQ metric did not improve its accuracy as much as using the equivalent SPD-NPS

parameters in other metrics. Using noise images and SPD-MTFs derived from dead leaves signals also did not improve its accuracy compared to using a uniform patch noise image and the direct dead leaves MTF. Implementing the cCSF and cVPF models reduced its accuracy greatly, probably due to the inhibiting effect of its curve fitting. Retuning the CPIQ metric with respect to the SPD-NPS, SPD-MTF and cCSF/cVPF parameters would improve the accuracy of metric variants that use them. It is first recommended that the CPIQ metric is tested with respect to a range of arbitrary MTFs, NPSs and CSFs of different shapes, to investigate its sensitivity to changes in each input parameter. Its sensitivity to changes in the shape of the CSF should be of particular interest, which was severe.

The cCSF, cVPF and other CSF models were cascaded as weighting functions by all metrics to account for the priority of different signal frequencies to perceived image quality. This has not been problematic traditionally, and many successful metrics implement CSFs in this way. However, it should be noted that these visual models are threshold functions for spatial vision, not transfer functions. There is less justification for them to be cascaded with the image signal compared to transfer functions such as the MTF. The effect of higher-order processes of quality judgement on the priority of different signals to image quality are also unaccounted for by all these visual models. This should be investigated in further work.

The Barten or Johnson and Fairchild CSFs are commonly applied as weighting functions by IQMs. Normalising the cCSF and cVPF to the same integrated area as these models rendered them more suitable weighting functions. After this normalisation, the IQMs still accounted for scene-dependent differences in the shape of the cCSF and cVPF (shown in Figures I3 and I4). However, the IQMs did not account for variations in the magnitude of these functions which relate to the contrast of the scene. The revised SQRIn and PIC metrics still took scene contrast into consideration since they accounted for the power spectrum of the scene image. But this was not the case for the Visual log NEQ. Disabling the normalisation of the cCSF and cVPF did not improve the accuracy of any of the IQMs (Tables J2 to J6).

Empirical data and discussions from prior art suggest that the cCSF and cVPF are more suitable visual functions for image quality modelling than traditional CSFs (as detailed in Section 2.5). The fact that metrics using the cCSF and cVPF were less accurate than metrics using the Barten CSF was unexpected. For the SQRIn, PIC and CPIQ metrics, it is assumed this was caused by their calibration being disturbed when the cCSF/cVPF were implemented. However, this does not explain the loss of accuracy in the Visual log NEQ metric that applies

minimal calibration. Potential explanations for why the cCSF and cVPF reduced metric accuracy are given below.

The cCSF and cVPF account for visual masking using the Linear Amplification Model (LAM). The LAM models the effect of masking on detection/discrimination of signals in a given frequency band according to the contrast of flanking bands in the contrast spectrum. This global model does not account for whether or not the signal being detected/discriminated is located next to any masking content. If a local masking model were used instead, detection/discrimination would be modelled as varying locally (as in practice) dependent on the proximity of the signal to masking content. It is expected that the cCSF and cVPF would be implemented more successfully in IQMs if they applied such a model. However, it would make them significantly more complex.

7.5 Implementation and Application

7.5.1 Practical Implementation of SPD-MTF and SPD-NPS measures

It is proposed that the SPD-MTF and SPD-NPS measures developed in this thesis are implemented with real capture systems by one of the following two methods.

The first adapts the implementation of Branca *et al.* [7], as summarised below:

- 1) Capturing the test scenes/target with a very-high-resolution, professional camera ensuring these captures contain no visible artefacts (Branca *et al.* [7] used a Hasselblad 503 camera with a 28 megapixel Leaf Aptus digital back and standard 80mm f/2.8 lens).
- 2) Printing the test scenes/target at the largest scannable size using a professional printer with a very high dynamic range and resolution (Branca *et al.* [7] used an Epson Stylus Pro 7900 printer with resolution of 2880 x 1440 dots per inch (dpi)).
- 3) Mounting the printed images flat and positioning them parallel to the sensor of the test device at a distance that results in a sampling density of 140 pixels per inch (ppi) or lower when they are captured. Illuminating the printed images evenly. Note that higher sampling densities may introduce measurement error.

-
- 4) Capturing the printed images, including any replicates, with the test device mounted on a tripod. A cable shutter release should be used, and the test device's mirror lock-up feature should be enabled, if available.
 - 5) Scanning each printed image using a professional scanner with a pixel resolution of 1200 dpi or higher (Branca *et al.* [7] used the Epson Perfection V850 Pro scanner). The MTF and NPS of the scanner should be negligible compared to the test device under the capture conditions. Otherwise, the ISO 12233 slanted edge MTF and ISO 15739 [168] uniform patch NPS of the scanner can be obtained and accounted for in 6). The scanner should be linear to ensure these measurements are representative of its average real-world performance.
 - 6) Resizing the scanned images to the same pixel resolution as the test device captures, using *lanczos3* interpolation [343]. Computing their one-dimensional (1D) discrete Fourier transform (DFT) power spectra (Equation 2.30). Compensating for the MTF and NPS of the scanner, if necessary, to yield suitably scaled 1D DFT power spectra for each print presented to the test device.
 - 7) Computing the desired SPD-NPS for the device from the replicate captures of 4).
 - 8) Calculating the SPD-MTF of the device with respect to the captured scene/target, scanned scene/target, and SPD-NPS (given by 4), 6) and 7), respectively).

Note that the SPD-NPS measures can also be computed from registered captures of real static scenes, by adapting step 4) accordingly, and then following step 7).

The second method requires an automated camera characterisation system to be developed, illustrated in Figure 7.1. It enables SPD-MTFs and SPD-NPSs to be obtained conveniently from a large number of replicates, reducing error. It should also allow them to be computed in real-time from video signals, as suggested in further work.

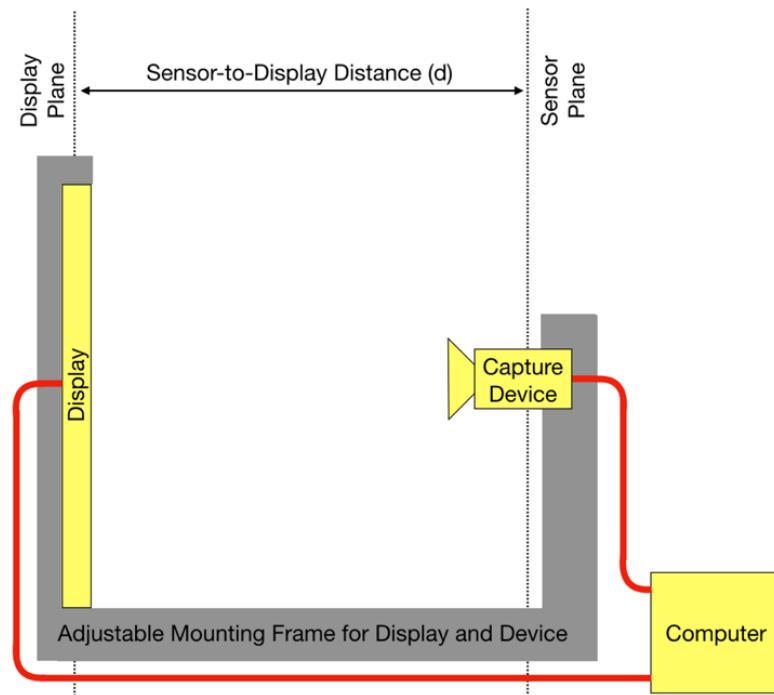


Figure 7.1 Proposed automated display-capture device layout.

The very-high-resolution display should deliver a linearised, uncompressed output signal. Its MTF should be characterised following reference [344] or [102, p. 208]. Its NPS should be characterised by subtracting the uniform patch NPS of a professional digital single-lens reflex (DSLR) camera (measured following ISO 15739 [168]) from the combined NPS of the display and DSLR. The latter should be measured by capturing a similar uniform tone signal with the DSLR, presented by the display. Incorrect characterisation of the display will result in measurement error; the DSLR must not apply non-linear ISP. The mounting frame should align the display and sensor planes precisely to mitigate aliasing and minimise vibration of either component. The sensor-to-display distance should be sufficient to ensure the pixel resolution and MTF of the display both outresolve the capture device.

Bespoke software should display the images of the test scenes/target and trigger the device to capture them, automatically, including any replicates. It should then compute the SPD-NPS of the device, by subtracting the NPS of the display from the SPD-NPS of the captured scene/target images. The SPD-MTF of the device should then be computed with respect to its SPD-NPS, and the 1D DFT power spectra of the captured and displayed images. The displayed image should be resized to the same pixel resolution as the captured image using *lanczos3* interpolation [343], before computing its power spectrum. Its power spectrum

should also be multiplied with the square of the MTF of the display to account for the display's limitations.

7.5.2 Application in Objective Imaging Performance Metrics

Objective metrics for imaging system noise and its scene-dependency have been developed by the author in conjunction with this project [345].

The *mean pictorial image SPD-NPS area* metric, A_{Mean} , describes the objective level of temporally varying system noise as a single figure (Equation 7.1) [345]; it accounts for system noise scene-dependency. $N(u)$ is the mean pictorial image SPD-NPS. u and $u_{Nyquist}$ are spatial frequency and the Nyquist frequency, respectively. Like the MTF50 sharpness metric and the MTF10 resolution metric, this objective metric for system noise does not account for the human visual system (HVS). Thus, it should not be confused with IQMs and should be applied as a system noise optimisation parameter.

$$A_{Mean} = \int_0^{u_{Nyquist}} N(u).du \quad (7.1)$$

The *relative standard deviation area (RSDA) of the pictorial image SPD-NPS*, A_{RSDA} , expresses the objective level of scene-dependency of temporally varying system noise [345]; it also does not account for the HVS. It is given by Equation 7.2, where $S(u)$ is the pictorial image SPD-NPS standard deviation (Equation 4.7); other parameters are as above.

$$A_{RSDA} = \frac{\int_0^{u_{Nyquist}} S(u).du}{A_{Mean}} \quad (7.2)$$

Both metrics were validated in reference [345] with respect to images generated by pipelines tuned at full (Chapters 4 and 5) and reduced opacity (Chapter 6). The mean pictorial image SPD-NPS area metric scores (Figure 7.2) corroborated observations of mean pictorial image SPD-NPS measurements of system noise performance, including those from Chapter 4. For example, the former indicated that non-linear denoising and sharpening algorithms removed noise more and amplified noise less, than the equivalent linear algorithms, respectively.

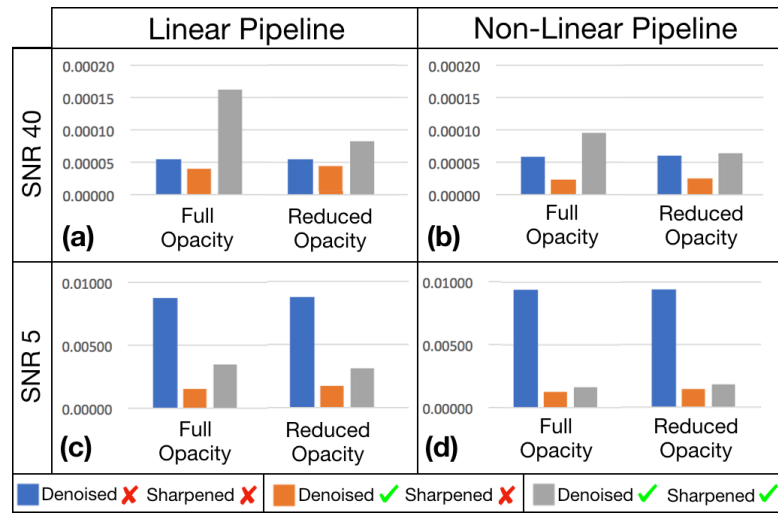


Figure 7.2 Mean pictorial image SPD-NPS area, in units of pixels.

Likewise, the RSDA metric (Figure 7.3) agreed with observations of pictorial image SPD-NPS standard deviation measurements of system scene-dependency. It was unaffected by linear ISPs, as would be expected. Non-linear denoising increased it significantly, but not non-linear sharpening. The latter did not increase the relative spread of the pictorial image SPD-NPS curves (Figure 4.7) but did re-arrange their shape and order. Both the last are unaccounted for by the RSDA metric but affect system performance scene-dependency.

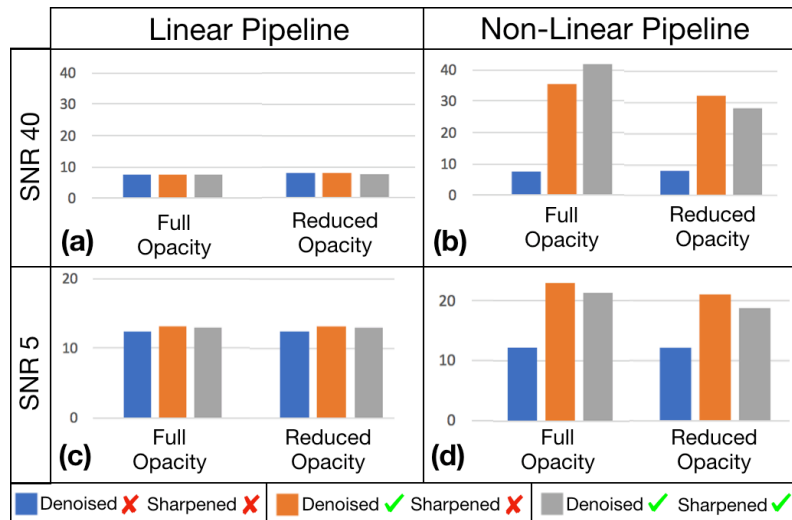


Figure 7.3 Relative Standard Deviation Area of the SPD-NPSs of 50 input scenes. This is expressed as a percentage of the integrated area under the mean pictorial image SPD-NPS.

Quoting the RSDA metric alongside the mean pictorial image SPD-NPS area metric expresses both the systems performance and scene-dependency, in the same manner as when the pictorial image SPD-NPS standard deviation is added/subtracted from the mean pictorial

image SPD-NPS in Figure 4.6. This is useful when assessing and benchmarking system performance. For example, systems with a low mean pictorial image SPD-NPS area metric score, and a high RSDA score are likely to use significant non-linear ISP to improve their output image quality, as well as good quality hardware.

Substituting the equivalent SPD-MTF measures into Equations 7.1 and 7.2 yields objective metrics for system signal transfer (accounting for system scene-dependency) and its scene-dependency, respectively. Both these metrics are currently limited by bias in the pictorial image SPD-MTF (discussed in Chapter 5) and would benefit from it being mitigated further.

7.5.3 Application in Computer Vision and Autonomous Vehicles

The performance measures and metrics proposed in this thesis are suitable for a wide variety of applications beyond visual image quality modelling. For example, the SPD-MTF, SPD-NPS and scene-and-process-dependent NEQ (SPD-NEQ) can be used to characterise capture systems used in computer vision applications. This includes systems that deliver input video streams to the deep Convolutional Neural Networks (CNN) used by Advanced Driver Assistance Systems (ADAS). These CNNs detect and classify road signs and hazards. The video stream is usually processed by a system on chip (SoC) before being analysed by them [346]–[349]. SoCs are generally black boxes that automatically apply non-linear content-aware ISP such as denoising, video noise temporal filtering (VNTF) [350] and sharpening [351]. These algorithms are dependent on the spatial or spatiotemporal signal characteristics.

CNNs are on par or better than human observers at classifying objects in high-quality images of scenes, but far less accurate for blurred, noisy images [352]. ISP algorithms should also be tuned differently in systems intended for CNNs and human observers, to maximise classification accuracy [351], [353]. Jenkin and Kane [351] suggest CNNs should be treated presently as “alien observers” and the imaging industry is a long way from producing IQMs calibrated for CNNs in the manner of the psychovisually calibrated CPIQ metric. What is clear is that for the performance of ADAS systems and relevant metrics to be reliable, their capture systems and SoC pipelines must be characterised accurately.

Jenkin and Kane [351] have proposed a fundamental spatial signal model for autonomous vehicle capture systems. It is based on the Effective Pictorial Information Capacity (EPIC) metric [195], [252] that relates to the information capacity and NEQ (as discussed in Section 2.4). It is proposed in this thesis that the “photographic” expression for information capacity

(Equation 2.41) [173, p. 631] should be revised for the same purpose. The signal spectrum of the displayed image, $S(u)$, and the total imaging system noise, $N(u)$, should be obtained using Equations 3.15 and 3.16, respectively. $\gamma_{display}^2$ and $MTF_{display}^2(u)$ should be replaced with constants of value 1 since no display exists in this scenario. The resultant measure accounts for the SPD-MTF and SPD-NPS, and thus also capture system scene-dependency.

Kane [342] and Jenkin [354] also developed an Ideal Observer Signal-to-Noise Ratio (SNRI) model for autonomous vehicle capture system design and optimisation (Equation 7.3). It describes the SNR concerning the decision of the ideal observer, with respect to a given task, considering the SNR of the imaging system. It is based upon the detectability index, d' , that applies signal detection theory; both are discussed in Section 2.4. MTF_{sys} and NPS_{sys} are the 1D MTF and NPS of the imaging system, respectively. $G(u)$ is the DFT of the difference between the signal and background. K is the large area signal transfer factor.

$$SNRI^2 = K^2 \int \frac{|G(u)|^2 MTF_{sys}^2(u)}{NPS(u)} du = d'^2 \quad (7.3)$$

Jenkin states the SNRI can be obtained by cascading the NEQ with the target and background information and that reliable MTF and NPS measurements are critical to its accuracy [354]. The following *scene-and-process-dependent SNRI (SPD-SNRI)* measure is proposed in this thesis, in terms of the SPD-NEQ, $NEQ_{SPD}(u)$; all other parameters are as per Equation 7.3. Its similarity with the log NEQ should be noted. The fact the logarithm is not taken is understandable considering that, unlike the human visual system, CNNs do not necessarily follow Fechner's law.

$$SNRI_{SPD}^2 = K^2 \int |G(u)|^2 NEQ_{SPD}(u) du = d'^2 \quad (7.4)$$

Both of the proposed spatial signal models account for capture system signal transfer and noise scene-dependency caused by interactions between relevant input signals and non-linear ISP algorithms. The proposed models would benefit from further reduction of pictorial image SPD-MTF measurement error. Computing each model using the pictorial image SPD-MTF and SPD-NPS accounts for system performance with respect to real video frames. Using the mean pictorial image SPD-MTF and SPD-NPS instead, or the dead leaves SPD-MTF and SPD-NPS, accounts for the system's average real-world capabilities, including its scene-dependency.

Chapter 8 Conclusions and Recommendations for Further Work

8.1 Conclusions

The following conclusions were drawn from research in this thesis:

- The novel image quality metrics (IQM), *scene-and-process-dependent Modulation Transfer Functions (SPD-MTF)* and *Noise Power Spectra (SPD-NPS)* of this thesis represent a new paradigm of IQMs and imaging performance measures. They accounted for the scene-dependency of simulated non-linear image capture pipelines. But they are more complex than current equivalent measures.
- The *pictorial image SPD-MTF and SPD-NPS* are the only current measures that characterise non-linear system performance with respect to a given input scene; they accounted most thoroughly for system scene-dependency. They are the most suitable input parameters for IQMs when modelling the quality of a given captured scene.
- The *mean pictorial image SPD-MTF and SPD-NPS* are, presently, the only measures for the average real-world performance of non-linear systems that account for system scene-dependency. They are the most suitable IQM input parameters when describing the average real-world image quality of such systems.
- The *pictorial image SPD-MTF and SPD-NPS standard deviation* are the only current measures for the level of system scene-dependency. But they do not account for all aspects of it. Combining them with the mean pictorial image SPD-MTF and SPD-NPS characterises average real-world system performance and scene-dependency.
- The *dead leaves SPD-MTF and SPD-NPS* characterised the average real-world performance of the non-linear pipeline more appropriately than the direct dead leaves Modulation Transfer Function (MTF) and uniform patch Noise Power Spectrum (NPS), respectively. They are more convenient and computationally efficient than the mean pictorial image SPD-MTF and SPD-NPS and suit the same applications. But they were often outliers compared to measures derived from images of scenes,

since the dead leaves chart triggered non-linear image signal processes (ISP) at different levels to natural scene signals.

- All SPD-MTF measures suffered from bias due to signal-to-noise limitations. This bias was scene-dependent, affecting in particular the measurements at the higher frequencies of noisy, low-signal images of scenes. It was mitigated by denoising.
- Fixed patterns of noise and artefacts were unaccounted for by all SPD-NPS measures.
- The accuracy of the novel *log Noise Equivalent Quanta (log NEQ)* and *Visual log NEQ* metrics developed in this thesis, as well as other leading IQMs, improved when they used SPD-MTFs and SPD-NPSs as input parameters.
- The greatest improvements in accuracy occurred when the metrics used SPD-NPS measures derived from images of scenes. This shows the SPD-NPS framework is robust and relevant to image quality modelling.
- When the metrics used SPD-MTFs measured from images of scenes, the benefits of accounting more thoroughly for system scene-dependency were often traded off against bias from signal-to-noise limitations.
- Implementing Contextual Contrast Sensitivity Function (cCSF) and Visual Perception Function (cVPF) spatial visual models did not improve metric accuracy.
- The IEEE P1858 Camera Phone Image Quality (CPIQ) standard metric [22] achieved the highest accuracy of all metrics tested when specific input parameter combinations were used (namely, the Barten CSF and SPD-MTFs and noise images derived from images of scenes). Other input parameter permutations, especially those involving the cCSF and cVPF, appeared to disturb the metric's calibration and reduced its accuracy.
- The log NEQ and Visual log NEQ applied minimal calibration. Thus, analysing changes in the accuracy of these metrics informed most of all on the appropriateness of the various input parameters. These metrics were generally more accurate than the comparable Pictorial Information Capacity (PIC) and Square Root Integral with noise (SQRI_n) and more consistent than the CPIQ metric. This demonstrates the relevance of the Noise Equivalent Quanta (NEQ) and novel *scene-and-process-dependent NEQ (SPD-NEQ)* measures to image quality modelling. It also suggests simple, elegant metrics may be the most accurate, provided their input parameters characterise system performance comprehensively. This is in agreement with Occam's Razor.

8.2 Further Work

The following further work is recommended that relates to the subjects covered in this thesis:

- Validation of all measures and metrics of this thesis with real capture systems using the practical implementations proposed in Section 7.5.1. This should include systems applying Joint Photographic Experts Group (JPEG) and JPEG 2000 compression.
- Evaluation of variation error and bias in the SPD-MTF and SPD-NPS measures from first principles. This can be achieved for the SPD-MTFs by modifying Burns' error propagation method for existing dead leaves MTF implementations [158].
- Investigation of methods to mitigate bias in the SPD-MTFs resulting from signal-to-noise limitations.
- Validation of all measures and metrics of this thesis at higher noise levels. This would require extending the low-quality range of the ISO 20462 image quality ruler.
- Investigation of whether computing the SPD-MTF and SPD-NPS of video frames accounts for the spatiotemporal-signal-dependency of non-linear video processing.
- Comparison of the dead leaves SPD-NPS and Artmann's [18] noise measures derived from the dead leaves test chart.
- Analysis of the effect of common Colour Correction Matrix (CCM) on the SPD-NPS measurements.
- Analysis of SPD-MTF and SPD-NPS measurements with respect to local regions of images of scenes, and how they integrate into the global SPD-MTFs and SPD-NPSs of this thesis. Such local measurements could be implemented in IQMs alongside visual saliency models, or in signal detection metrics for autonomous vehicles.
- Integration of envelopes of scene-edge Spatial Frequency Response (SFR) measurements (Section 7.3) into global SFRs [341]. These measurements should be compared with SPD-MTFs, including their use as IQM input parameters.
- Validation of the signal models for autonomous vehicles proposed in Section 7.5.3.
- Examination of the relationships between natural scene statistics (NSS) and typical SPD-MTF, SPD-NPS and SPD-NEQ behaviour, as well as bias in each measure.
- Investigation of the role of masking scene content to perceptions of sharpness and noisiness, as well as denoising, sharpening and compression artefact magnitude.

- Evaluation of whether contextual detection (cCSF) and discrimination (cVPF) models have genuine roles to play in image quality modelling.
- Implementation of optimal contrast weighting (OCW) functions [14] in the IQMs of this thesis to account for relevant cognitive processes of image quality judgement.
- Benchmarking the IQMs of this thesis versus state-of-the-art Computational IQMs (CP-IQM) using the observer image quality rating dataset from Chapter 6.
- Generation of a test image dataset which triggers a range of non-linear spatial ISP algorithm behaviour in terms of signal transfer (SPD-MTF) and noise (SPD-NPS). Susceptibility to sharpening and denoising should be considered.
- Development of a test chart that: 1) relates to the power spectrum, spatial distribution of contrast signals and range of edge gradients of the “average scene”. 2) is shift, rotation and exposure invariant, like the dead leaves chart. The dead leaves SPD-MTF and SPD-NPS should be validated with the dead leaves chart substituted for it.

Appendix A Linear System Theory Requirements

All imaging system performance measures that apply linear system theory require the system to be linear, spatially invariant and homogenous. This applies to the Modulation Transfer Function (MTF), Noise Power Spectrum (NPS) and Noise Equivalent Quanta (NEQ). Each requirement is defined below, referring to Dainty and Shaw [6, pp. 206–207]. Failure to comply with them or compensate for any deviation results in measurement inaccuracy.

Let the operator $S\{ \}$ denote a 2D imaging system. The condition for the system to be linear is given below, where $f_1(x, y)$ and $f_2(x, y)$ are two separable input image signals [2, p. 4].

$$S\{f_1(x, y) + f_2(x, y)\} = S\{f_1(x, y)\} + S\{f_2(x, y)\} \quad (\text{A1})$$

The condition for the system to be homogenous is below; a is a scaling constant [2, p. 4].

$$S\{ax, ay\} = aS\{x, y\}; \quad \text{for } a > 0 \quad (\text{A2})$$

Combining Equations A1 and A2 gives the conditions for the system to be both linear and homogenous; a and b are constants greater than zero [6, p. 206].

$$S\{af_1(x, y) + bf_2(x, y)\} = aS\{f_1(x, y)\} + bS\{f_2(x, y)\} \quad (\text{A3})$$

The condition for the system to be stationary is given below [6, p. 207]. $h(x, y; x_1, y_1)$ is the response of the system at output coordinates x and y to an input signal at coordinates x_1 and y_1 . For a stationary system, this response depends on the differences between the variables only, $(x - x_1, y - y_1)$, not on each separate variable. Thus, the Point Spread Function (PSF) remains constant at all spatial locations of the image [6, p. 207].

$$h(x - x_1, y - y_1) \equiv h(x, y; x_1, y_1) \quad (\text{A4})$$

Appendix B Structural Similarity Index (SSIM)

Definition

Equation B1 [114] describes the luminance comparison of the Structural Similarity Index (SSIM), $l(x, y)$, with respect to a nonnegative reference, x , and test image signal, y . L is the pixel value dynamic range (i.e. 255 for an 8-bit grayscale image). K_{SSIM_1} is a constant of default value 0.01. Equation B2 defines the mean luminance of the reference image signal, μ_x , where x_i is the luminance at pixel i , and N is the total number of pixels [114]. μ_y is calculated from the test image signal, y , in the same fashion.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}; \quad \text{where } C_1 = (K_{SSIM_1}L)^2 \quad (\text{B1})$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (\text{B2})$$

The contrast comparison of the SSIM, $c(x, y)$, is a function of the standard deviations of the reference and test image signals (denoted by σ_x and σ_y , respectively) within a local window (Equation B3); K_{SSIM_2} is set to 0.03 by default [114]. σ_x is computed from the reference image signal, x , using Equation B4. σ_y is calculated from the test image signal in the same fashion. All other parameters are as detailed above [114].

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}; \quad \text{where } C_2 = (K_{SSIM_2}L)^2 \quad (\text{B3})$$

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (\text{B4})$$

The structural comparison, $s(x, y)$, is given as the correlation coefficient between the reference, x , and test image signal, y , respectively. It is computed by Equation B5 after the mean luminance, μ , has been subtracted from the signals, and they have been divided by their respective standard deviations, σ , to normalise them [114]. Thus, the structural

comparison describes the correlation between the unit vectors $(x - \mu_x)/\sigma_x$ and $(y - \mu_y)/\sigma_y$. σ_{xy} is estimated by Equation B6, where x_i and y_i are pixel intensities at position i for the reference and test images' signals, respectively, and μ_x and μ_y are the mean intensities of these signals [114]. N is the total number of pixels.

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}; \quad \text{where } C_3 = \frac{C_2}{2} \quad (\text{B5})$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (\text{B6})$$

The SSIM index between the reference, x , and test image signal, y , is given by Equation B7 where $\alpha = \beta = \gamma = 1$. This can be rewritten as Equation B8, where all terms are as described above [114].

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (\text{B7})$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (\text{B8})$$

A map of SSIM indices is output for a number of local windows of the image, M , that are produced using a circular-symmetric 11-by-11 pixel Gaussian weighting function. The SSIM map can be averaged to a single Mean Structural Similarity Index (MSSIM) figure using Equation B9. $SSIM(x_j, y_j)$ is the SSIM index corresponding to the reference and test image signals x_j and y_j , in the j th local window. X and Y are the reference and test images, respectively.

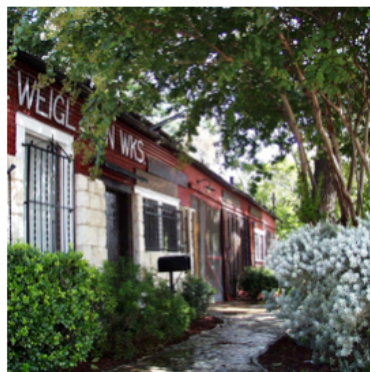
$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (\text{B9})$$

Appendix C Test Scenes used in Chapters 4 & 5

This appendix contains input scene images to the simulations used to validate the scene-and-process-dependent Noise Power Spectrum (SPD-NPS) and equivalent Modulation Transfer Function (SPD-MTF) measures in Chapters 4 and 5, respectively. Images 1-17 are from [225], 18-19 from [7], 20-23 from [14], 24 from [8], 25-26 from [102]. Images 1-13 from Appendix D were also used. Images 27-37 were captured with the same digital single-lens reflex (DSLR) camera and lens as images from [14] using identical focal length, aperture and ISO settings.



1. Bikes



2. Building2



3. Buildings



4. Caps



5. Coins In Fountain



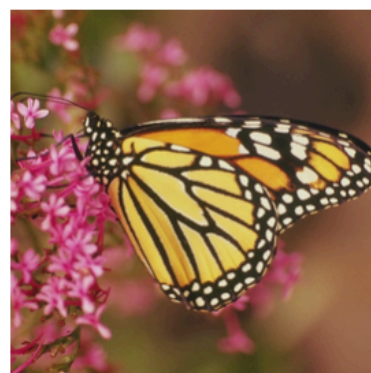
6. Flowers On IH35



7. House



8. Lighthouse II



9. Monarch



10. Ocean



11. Painted House



12. Parrots



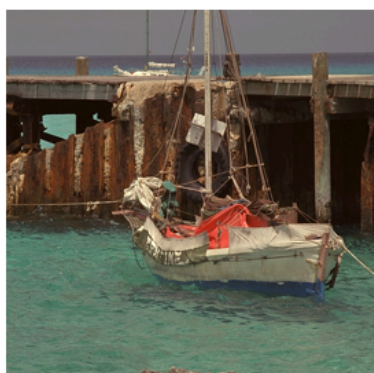
13. Plane



14. Rapids



15. Sailing I



16. Sailing IV



17. Stream



18. Architecture



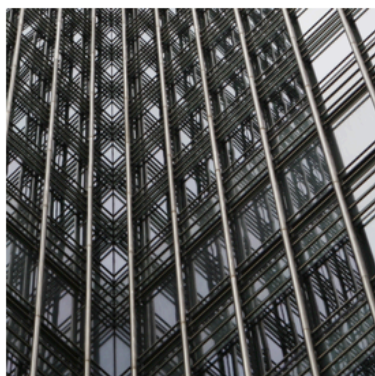
19. Students



20. Bench



21. People II



22. Lines



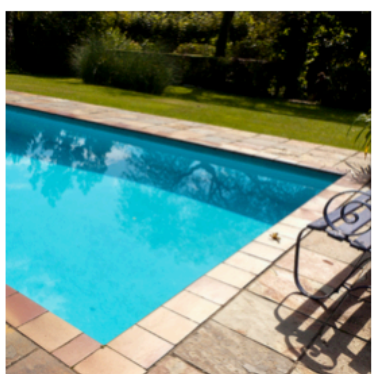
23. Buildings



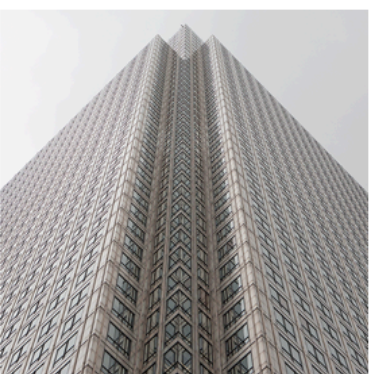
24. Gallery



25. Player's Navy



26. Pool



27. Architecture II



28. Landscape I



29. Landscape II



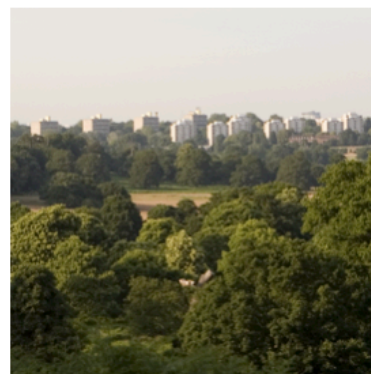
30. Landscape III



31. Landscape IV



32. Landscape V



33. Landscape VI



34. People II



35. People III



36. Panther



37. Ferry

Appendix D Test Scenes used in Chapter 6

This appendix contains input scene images to the simulations used to validate the revised image quality metrics (IQM) in Chapter 6. All were captured and processed by Allen [102].



1. Accordion



2. Afternoon Tea



3. Cliffs



4. Crockery



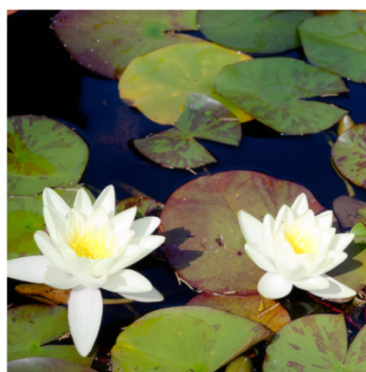
5. Emporium



6. Kids



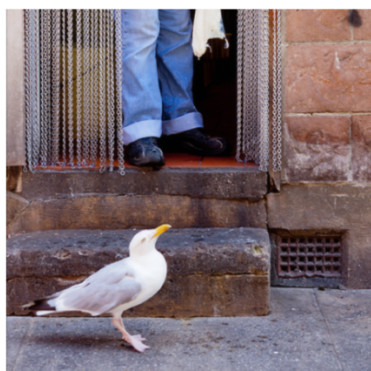
7. Lamp



8. Lilies



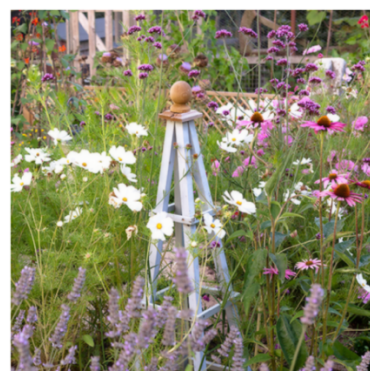
9. Sculpture



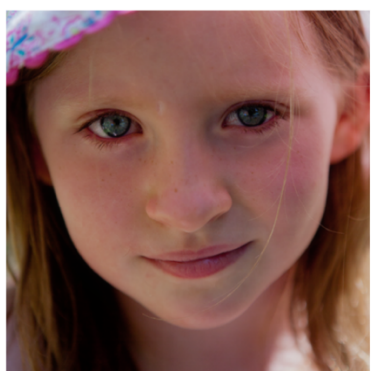
10. Seagull



11. Serpent



12. Flower Garden



13. Summer



14. Stones

Appendix E Generation of Ruler Images by Allen

Allen's [102, pp. 216–223] method for generating the ruler image dataset is summarised here. It followed Clause 7 of ISO 20462 [56, p. 9]. The ruler images were specific to the EIZO ColorEdge CG245W liquid crystal display (LCD) at 60cm viewing distance. They were generated from 14 captured scenes similar to those in Appendix D [102]. These original images of the scenes were of very high quality with no noticeable artefacts. But they differed in terms of subject matter and objective signal content. They were captured using a Canon EOS 5D Mark II digital single-lens reflex (DSLR) camera and Canon EF 24-70mm L II USM lens, using low ISOs and various focal lengths and F numbers [102, Sec. 5.2.1]. All images of the scenes were processed according to a standard imaging workflow [102, Sec. 5.2.2]. Cascading the Modulation Transfer Function (MTF) after processing with the MTF of the display yielded the system MTF (Figure E1(a)).

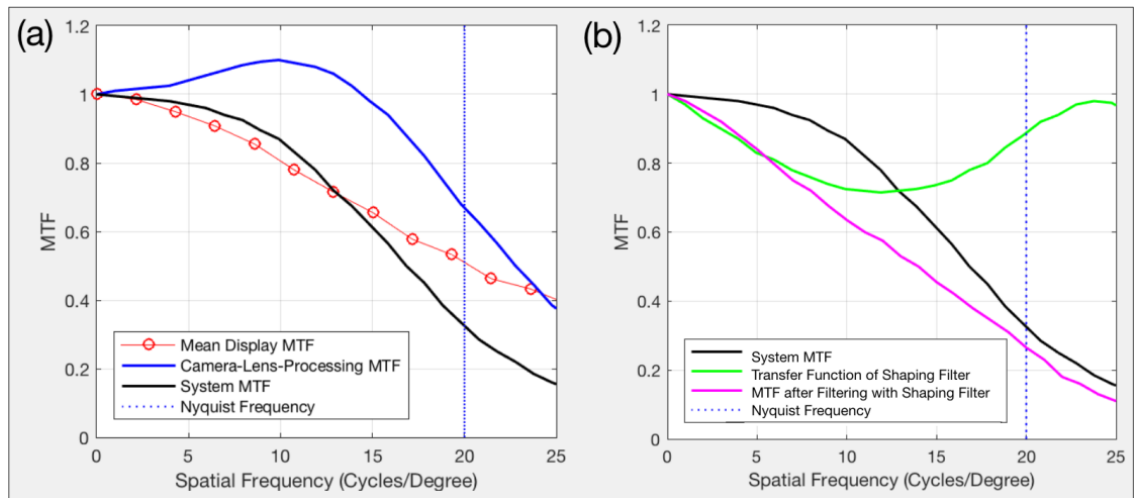


Figure E1 MTFs and filter transfer functions used by Allen [102, pp. 216–223] to shape the system MTF to the aim MTF.

ISO 20462 [56, p. 9] provides equations to generate ruler image MTFs (Equation E1) corresponding to a given Standard Quality Scale (SQS_2) value expressed in just-noticeable difference (JND) units (Equation E2). Both are given in terms of the constant k_{SCQR} ; u is spatial frequency. Allen [102, p. 91] determined an appropriate aim MTF for the system when k_{SCQR} was equal to 0.031, ($SQS_2 \approx 26$).

$$MTF(u) = \frac{2}{\pi} \cdot \left(\cos^{-1}(k_{SCQR}u) - k_{SCQR}u \sqrt{1 - (k_{SCQR}u)^2} \right) \quad k_{SCQR}u \leq 1 \quad (E1)$$

$$MTF(u) = 0 \quad k_{SCQR}u > 1$$

$$SQS_2 = \frac{17249 + 203792k_{SCQR} - 114950k_{SCQR}^2 - 3571075k_{SCQR}^3}{578 - 1304k_{SCQR} + 357372k_{SCQR}^2} \quad (1 \leq 100k_{SCQR} \leq 26) \quad (E2)$$

Allen [102, p. 217] divided the aim MTF by the system MTF to obtain the transfer function for a frequency domain shaping filter (Figure E1(b)). A 5th degree polynomial curve was fitted to this transfer function.

The scene image's MTF was then shaped to the aim MTF by [102, p. 217]:

- 1) padding the image with zeros to avoid wraparound error;
- 2) generating an array of ones, of identical dimensions to the padded image from 1);
- 3) applying the abovementioned polynomial function radially to the array from 2), with zero frequencies at the centre of the array;
- 4) computing and centring the two-dimensional (2D) Fast Fourier Transform (FFT) for the image to be filtered;
- 5) multiplying the magnitude component of the FFT from 4) with the filter array from 3);
- 6) recombining the result of 5) with the phase component of the FFT from 4);
- 7) inverse FFT filtering and removing the zero-padding applied in 1).

Allen [102, p. 221] determined k_{SCQR} values for 30 ruler images using Equation E2. Their SQS_2 values differed from the SQS_2 value of the aim MTF by integer JND units, from +6 JNDs to -24 JNDs (i.e. $SQS_2 = 3$ to $SQS_2 = 32$). The required MTF for each ruler image was computed using Equation E1. Dividing this required MTF, by the system MTF, yielded a frequency domain JND filter transfer function for each ruler image, to which a 5th degree polynomial function was fitted. Up to 30 ruler images were produced for each scene by repeating step 1) to 7) with each polynomial function.

Appendix F Display Characterisation

The EIZO ColorEdge CG245W display (with specifications given in Table F1) was characterised by measuring the following:

- 1) white point and peak R, G and B channel output (Table F2);
- 2) positional non-uniformity (Figure F1);
- 3) tone transfer functions (Figure F2);
- 4) viewing angle dependency in terms of luminance and chrominance output (Figures F3 and F4);
- 5) short-term temporal stability (Figure F5).

The method of Park [355] was followed when measuring 2), 4) and 5).

	EIZO ColorEdge CG245W display
Display technology	Liquid crystal display (LCD)
Displayable area (cm)	51.8 (horizontal) x 32.4 (vertical)
Native pixel resolution (pixels)	1920 (horizontal) x 1200 (vertical)
Display colour	24bits (DVI) / 30bits (DP) from a palette of 48 bits
Viewing angle (degrees)	178 (horizontal), 178 (vertical)
Pixel pitch	0.27mm (horizontal), 0.27mm (vertical)
Maximum brightness	270 cd/m ²
Maximum brightness for experiments	120 cd/m ²
Colour representation	Standardised RGB (sRGB)
Target gamma	2.2
Target white point	D ₆₅

Table F1 Display specifications; adapted from Allen [102, p. 299].

Before taking measurements from the display, it was allowed to warm up for 1 hour and then calibrated using its native colorimeter according to the target settings in Table F2(a). When characterising the display, the input signal was delivered by a Lenovo ThinkPad X230 with Intel HD Graphics 4000. The output signal was measured using a Konica-Minolta CS-200 colorimeter using the most accurate “slow” or “very-slow” measurement durations, with the

measurement angle set to 0.2 degrees. All measurements were taken in total darkness at a distance of 150cm. The temperature of the laboratory was approximately 20 degrees Celsius.

(a) Target Settings for Calibration							
	x	y	X'	Y'	Z'	u'	v'
White Point	0.3127	0.329	114.05	120.00	130.69	0.198	0.468
Peak Red	0.64	0.33	49.48	25.51	2.32	0.451	0.523
Peak Green	0.3	0.6	42.91	85.82	14.30	0.125	0.563
Peak Blue	0.15	0.06	21.66	8.66	114.08	0.175	0.158

(b) Measurements Obtained After Calibration							
	x	y	X'	Y'	Z'	u'	v'
White Point	0.318	0.3354	112.27	118.41	122.36	0.199	0.472
Peak Red	0.6458	0.334	50.64	26.19	1.58	0.452	0.526
Peak Green	0.302	0.6152	41.41	84.36	11.35	0.124	0.566
Peak Blue	0.1511	0.0609	21.02	8.47	109.60	0.176	0.160

Table F2 (a) target CIE xyY chromaticity coordinates for calibrating the display, and equivalent CIE 1931 tristimulus values and 1976 [356] chromaticity coordinates. (b) measurements of the above taken from the display after calibration using the Minolta CS-200 colorimeter.

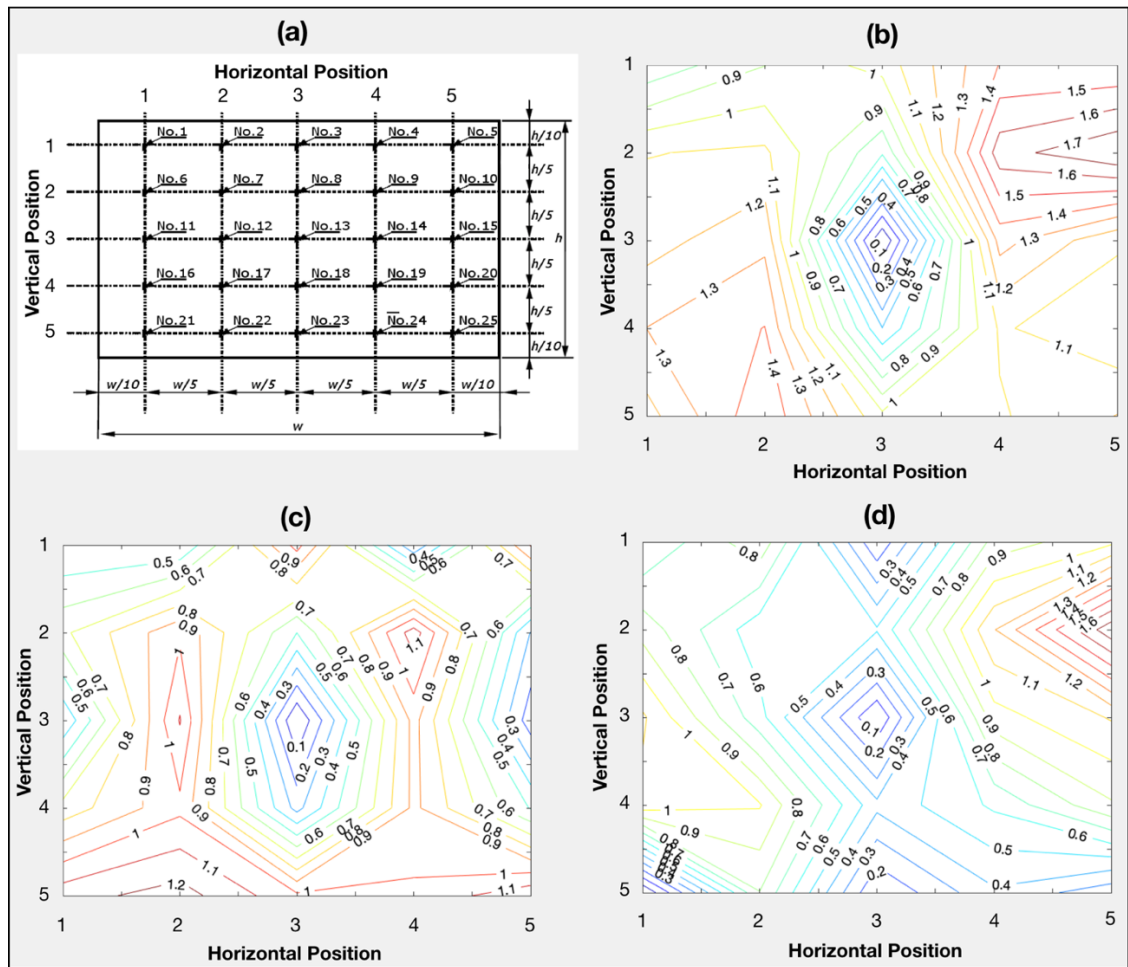


Figure F1 Measurements of the positional non-uniformity of the display: (a) shows the 25 positions where measurements were taken perpendicular to the display; h and w are the display's height and width [357, p. 25]. (b), (c) and (d) show CIE ΔE_{ab} colour differences, CIE ΔC_{ab} colour differences, and CIELAB ΔL^* lightness differences, respectively, between each display region and the centre of the display.

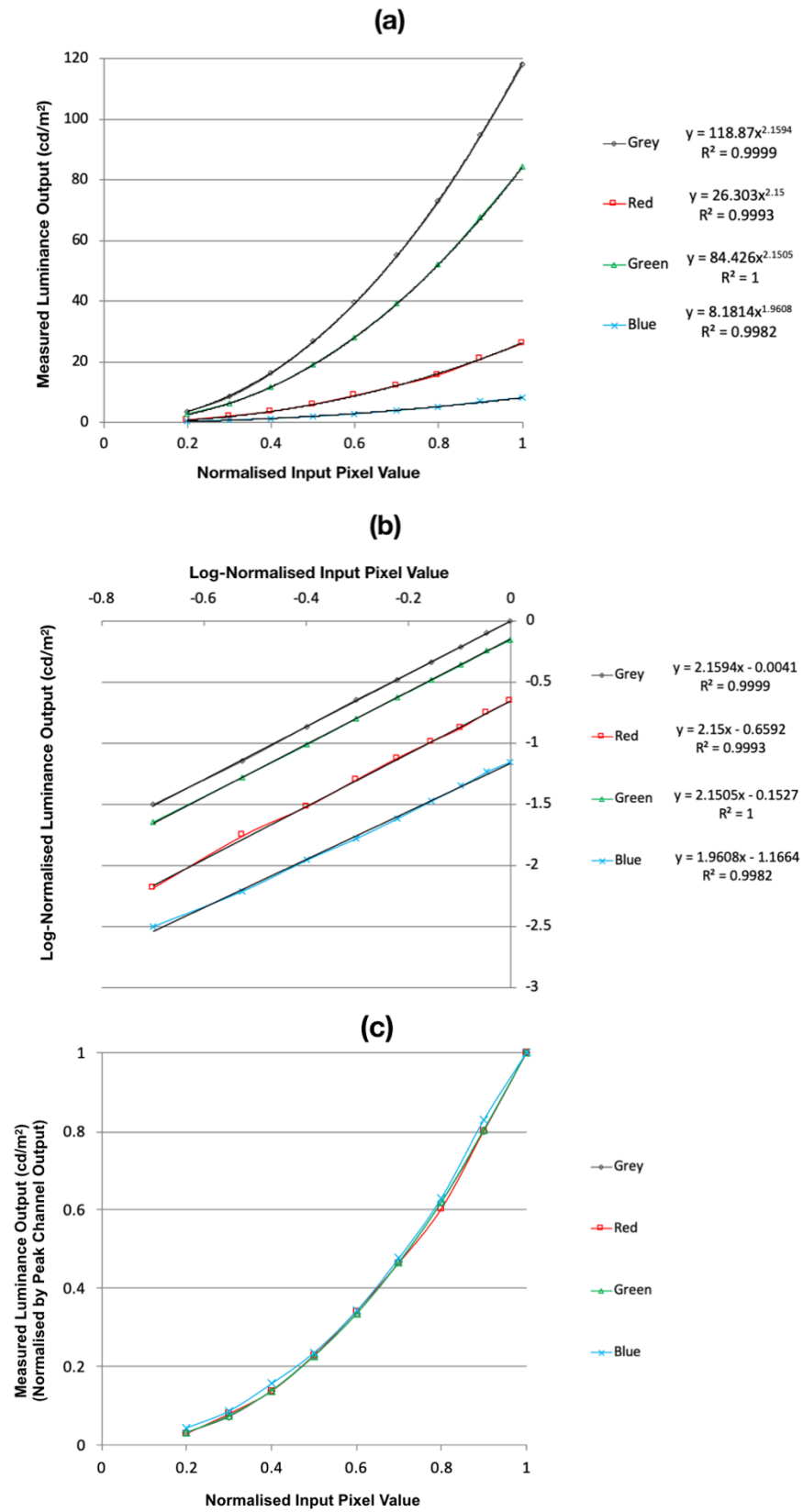


Figure F2 Tone transfer functions for the display: (a) linear-linear space, (b) log-log space, (c) normalised linear-linear space after dividing the output luminance of each channel by its peak luminance output. All measurements were taken perpendicular to the centre of the display. Measurements for 8-bit pixel values under 50 were omitted due to measurement inaccuracy.

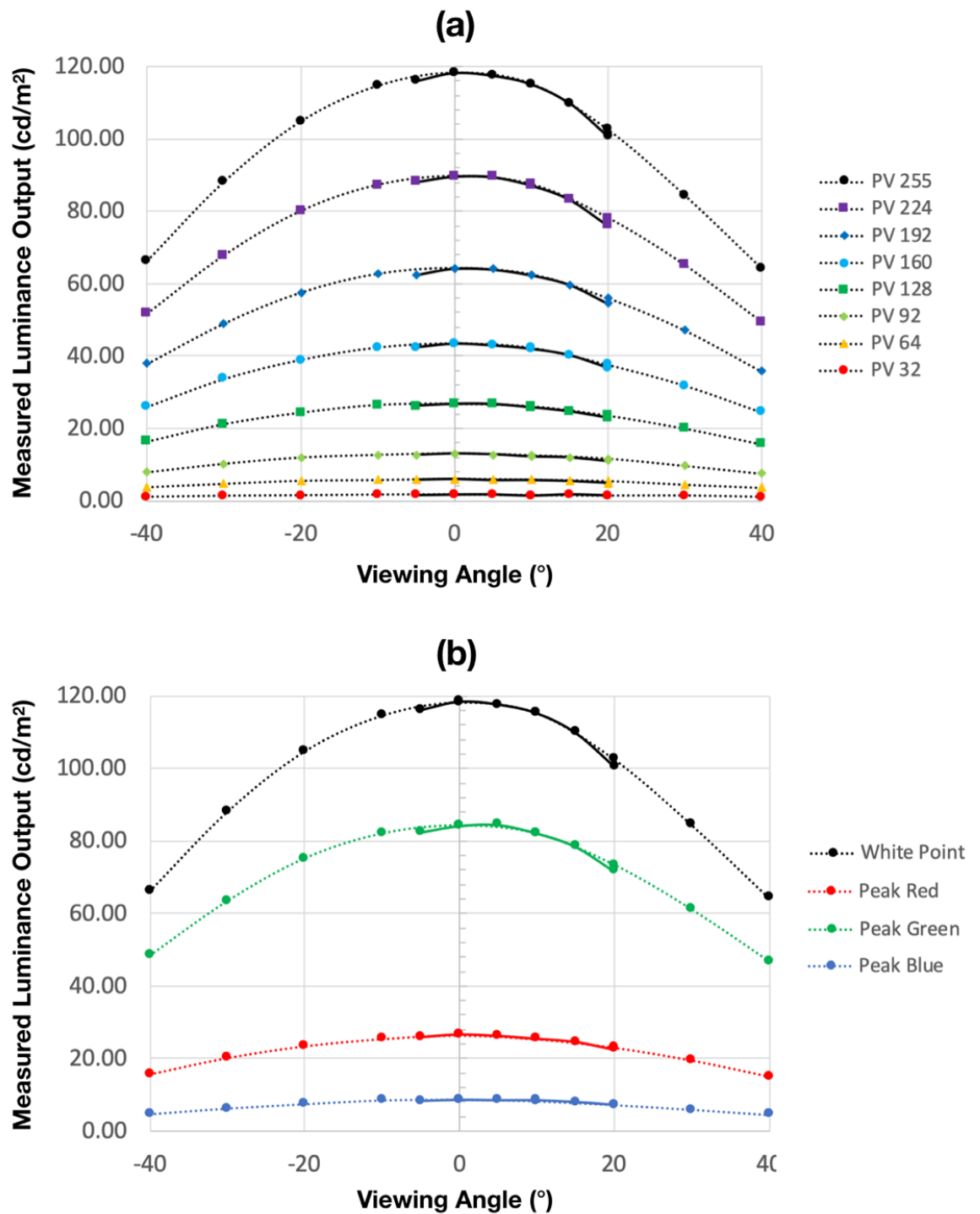


Figure F3 Luminance output dependency with respect to viewing angle for the display: (a) 8-bit neutral pixel values (PV) ranging from 32 to 255. (b) white point and peak output for the R, G and B channels. Horizontal and vertical viewing angle variations are denoted by dotted and solid trend-lines, respectively.

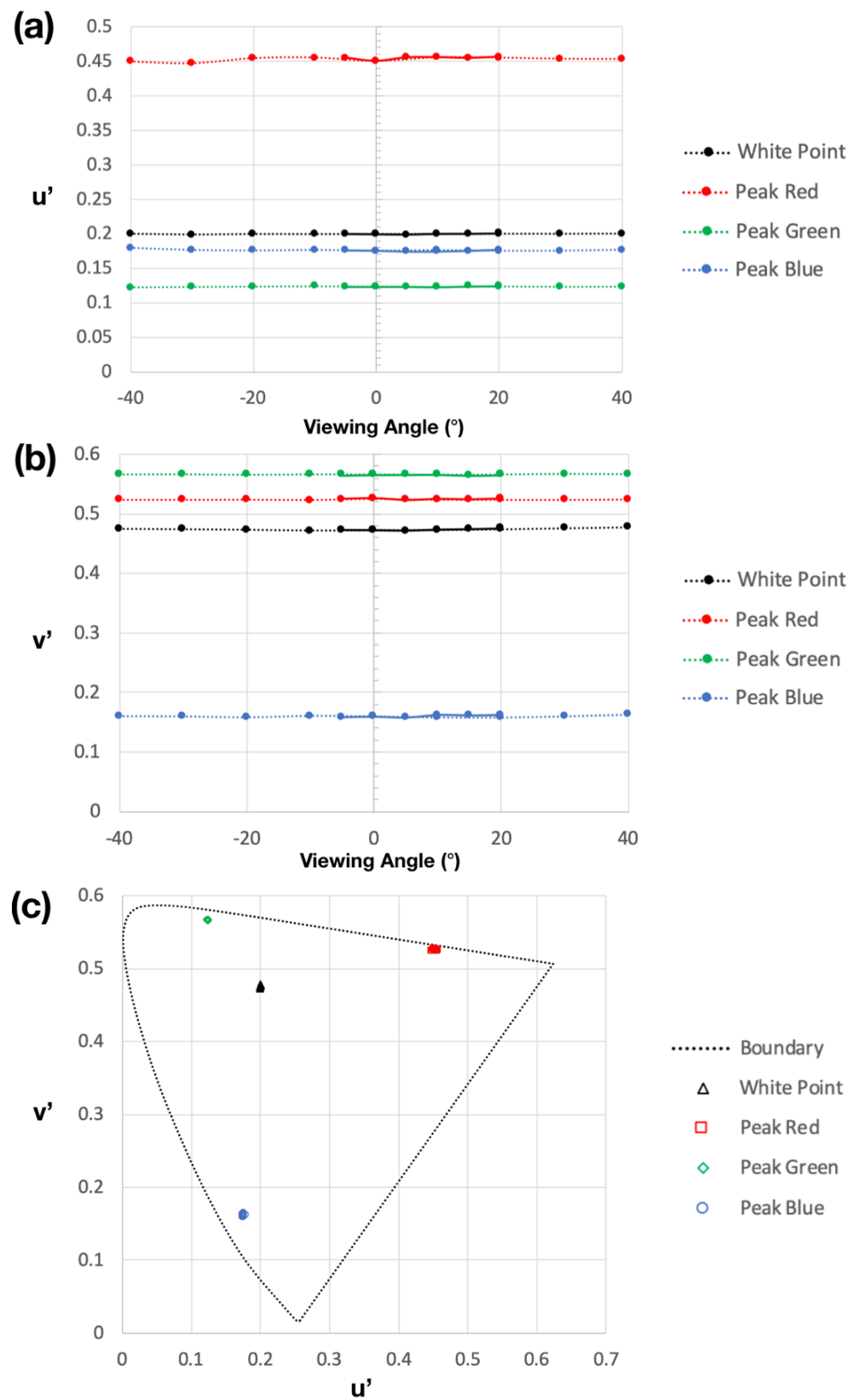


Figure F4 Variation in CIE 1976 [356] u' and v' chromaticity coordinates with respect to viewing angle for the display are shown in plots (a) and (b); measurements relating to horizontal and vertical viewing angle variations are denoted by dotted and solid trend-lines, respectively. (c) plots all data points from (a) and (b) on a standard CIE $u' v'$ chromaticity diagram.

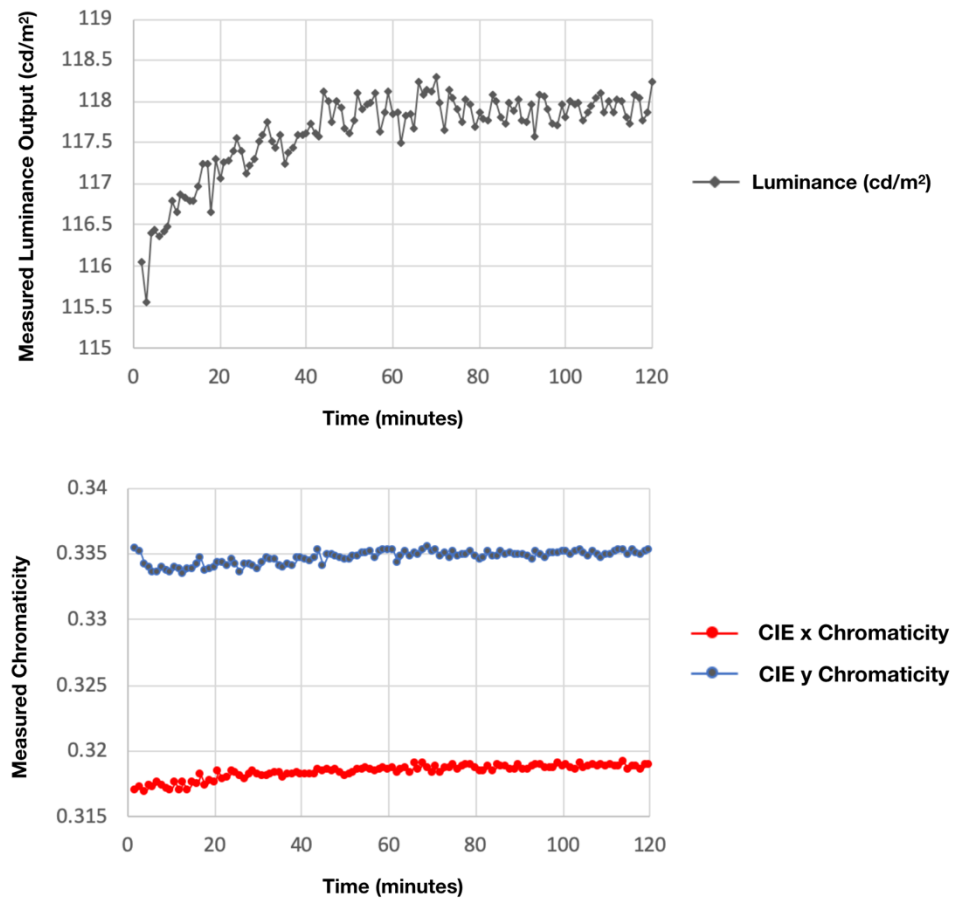


Figure F5 Temporal stability of the white point of the display with respect to luminance output (top), CIE 1931 x and y chromaticity coordinates (bottom).

Appendix G Snellen Near Vision Test Card

The Snellen [330] near vision test card (Figure G1) was printed at high quality using a laser printer. It was used to assess the near vision of participants in the image quality evaluations of Chapter 6.

Near Vision Test Card		
160 in.	E O P Z T L C D F	4.0 m
80 in.	T D P C F Z O E L	2.0 m
56 in.	D Z E L C F O T P	1.4 m
48 in.	F E P C T L O Z D	1.2 m
40 in.	P T L F C Z D E O	1.0 m
32 in.	E L Z T C O F P D	80 cm
24 in.	D Z E C L F T O P	60 cm
20 in.	L O P F E E D C T	50 cm
16 in.	E L T O F F D O E	40 cm

Figure G1 Snellen Near Vision Test Card [330].

Appendix H Observer Instructions

The following instructions were provided to each observer before their participation in the image quality evaluations of Chapter 6. They are based upon similar instructions from [56, p. 13] and [102, p. 306]. Numerical references in the text refer to sections of Figure H1.

You will be assessing the quality of a series of images using a psychophysical technique called the softcopy image quality ruler. Please remember that there are no right or wrong answers. Your opinions regarding image quality are a product of your own unique perceptions.

Here is how we are asking you to evaluate the images:

A pair of images will be presented on the display. The image on the left is labelled 'Ruler Image'. You can adjust its sharpness using the slider bar or arrow keys, to vary its quality. The image on the right is labelled 'Test Image' and will contain various artefacts. For each test image, we ask you to adjust the quality of the ruler image, until it matches the overall quality of the test image.

Please note that you will be balancing the quality lost due to blur in the ruler images, to the quality lost due to various artefacts in the test images. When you are comparing the test and ruler images, ask yourself which one you would keep if this were a treasured image and you were allowed to keep only one copy. If you prefer the test image, then you should move the slider bar to the left for a sharper ruler image. Alternatively, if you prefer the ruler image, then you should move the slider bar to the right for a more blurred ruler image. When you have finished adjusting the ruler, the two images should be of equal quality in your opinion. Your response will be recorded when you press the 'Next' button.

The evaluations will take between 1.5 and 2 hours in total and are separated into three stages. Stage 1 takes around 20 to 30 minutes. Stages 2 and 3 each take around 30 to 45 minutes. You should take a short break after each stage.

The test images were generated using a digital camera simulation. Test images from stages 1-3 correspond to different stages of in-camera image processing. Typical characteristics and artefacts for each stage are summarised below. Visual examples are provided on the next page, referenced numerically.

Stage 1 includes demosaicing (2) and noise (1) artefacts.

***Stage 2** includes demosaicing (2) and noise (1) artefacts. Denoising (3) is also applied. This reduces noise but causes blurring and/or loss of image texture and detail.*

***Stage 3** includes demosaicing (2) and noise (1) artefacts. Denoising (3) is applied (see Stage 2, above). Sharpening (4) is also applied that increases detail and edge intensity but also amplifies other artefacts.*

Before you participate in Stage 1, you will complete a brief trial run to familiarise yourself with the experiment. During the trial run, we encourage you to explain your decision making with the supervisor of the experiment. You can also ask questions at any time during the evaluations.

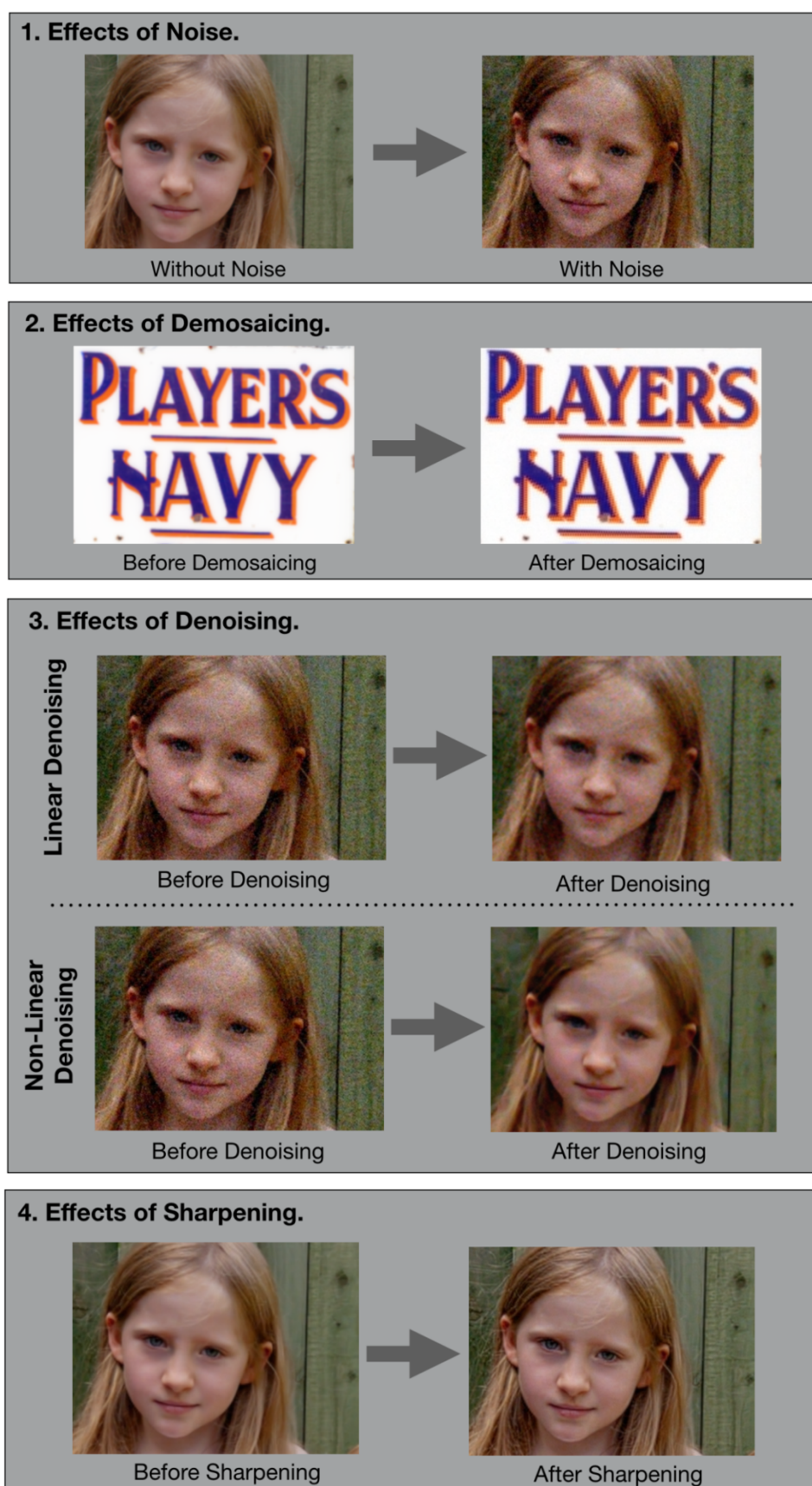


Figure H1 Examples of capture system artefacts, given to all participants with the instructions.

Appendix I IQM Input Parameters

The input parameters of all image quality metric (IQM) variants evaluated in Chapter 6 are presented. These include relevant imaging system performance measures and visual models.

Figures I1 and I2 present the Noise Power Spectrum (NPS) and Modulation Transfer Function (MTF) system performance measures, respectively. All measurements are from the simulated image capture pipelines defined in Section 4.2, where the opacity of the denoising and sharpening filters was reduced to levels given in Table 6.2 (refer to Chapter 6 for further details). Signal-to-noise ratios (SNR) of 80 and 10 correspond to the highest and lowest quality test images, respectively.

Figures I3 to I5 present various contrast sensitivity functions (CSF) used as input parameters to the IQMs (these functions are defined in Section 2.5). They are calculated with respect to output images from the mentioned pipelines.

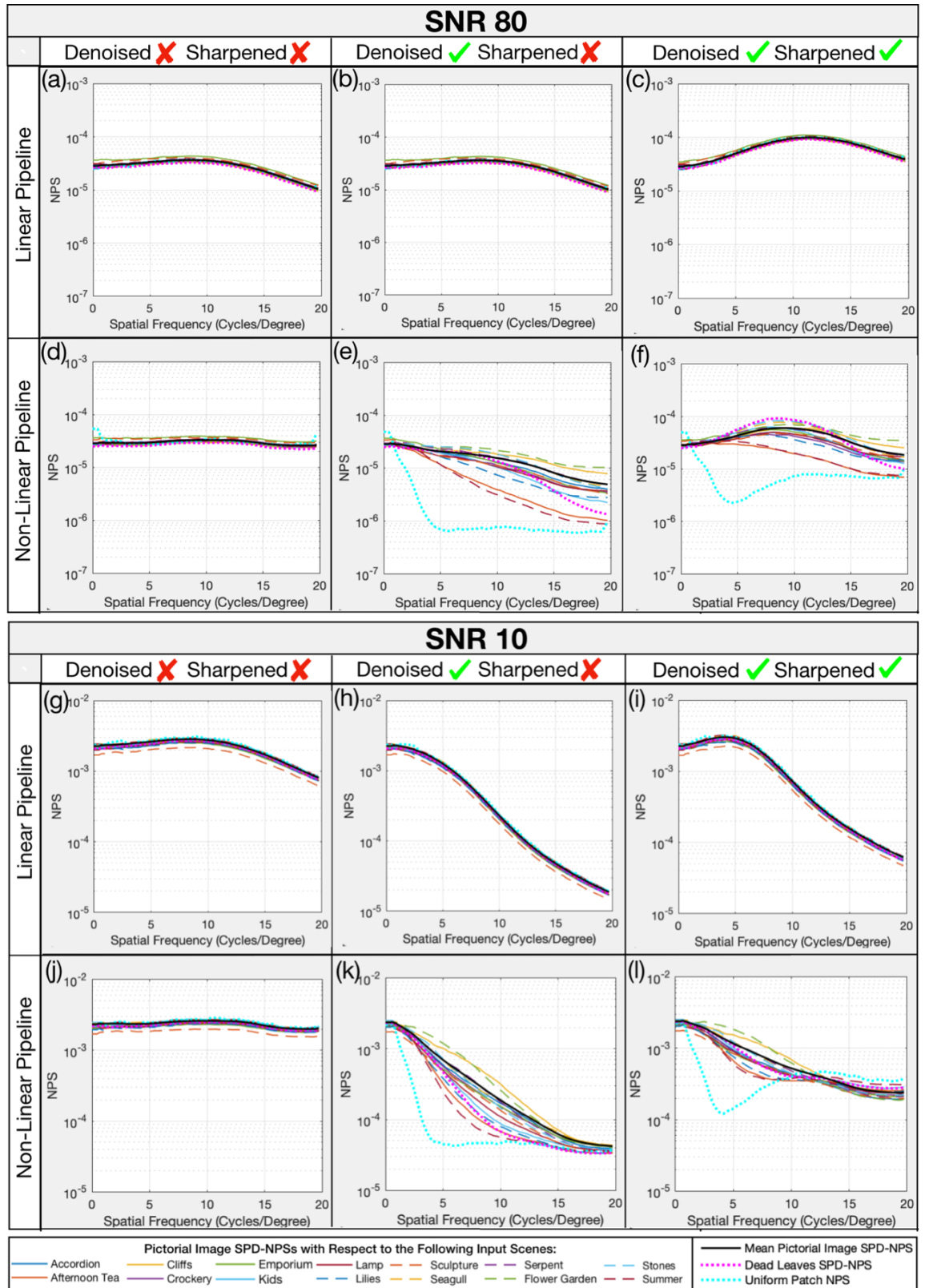


Figure I1 Noise Power Spectra (NPS) and scene-and-process-dependent NPS (SPD-NPS) measures used as input parameters to the image quality metric (IQM) variants in Chapter 6. Each measure is presented after relevant stages of linear and non-linear image signal processing (ISP).

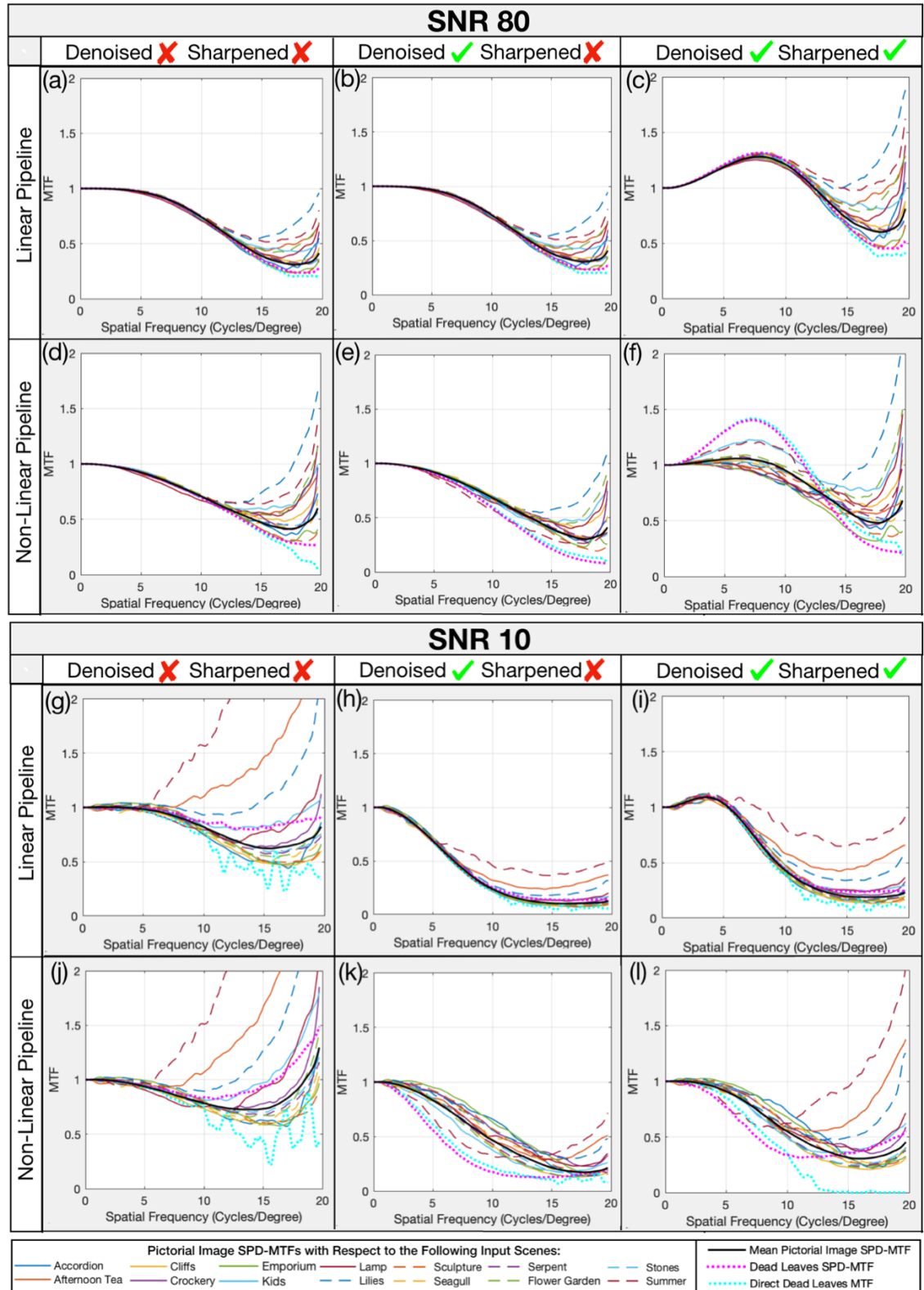


Figure I2 Modulation Transfer Function (MTF) and scene-and-process-dependent MTF (SPD-MTF) measures used as input parameters to the image quality metric (IQM) variants in Chapter 6. Each measure is presented after relevant stages of linear and non-linear image signal processing (ISP).

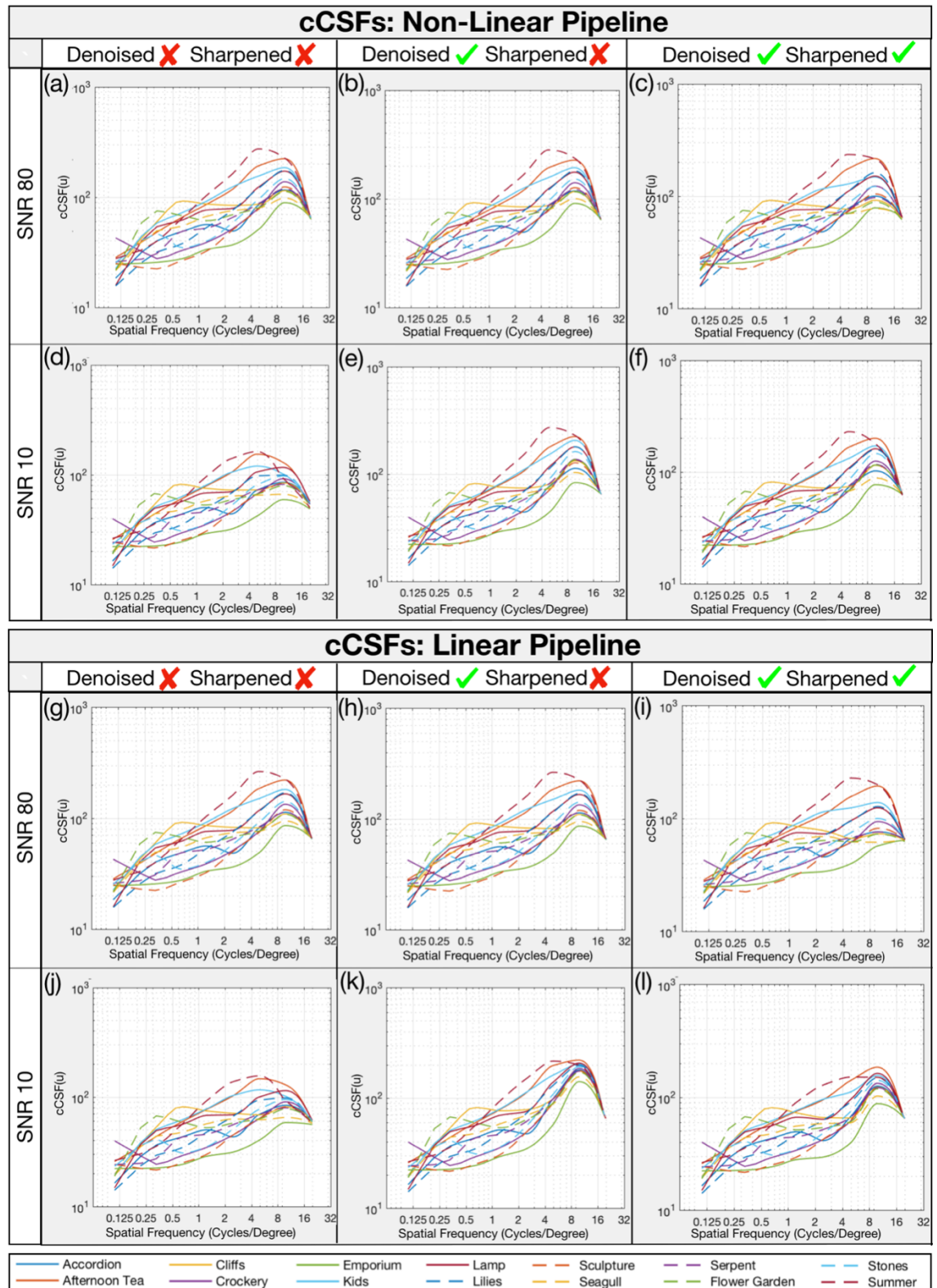


Figure I3 Contextual Contrast Sensitivity Functions (cCSF) [35] that were used as input parameters to the image quality metric (IQM) variants in Chapter 6. cCSFs for each input scene image are presented after relevant stages of linear and non-linear image signal processing (ISP).

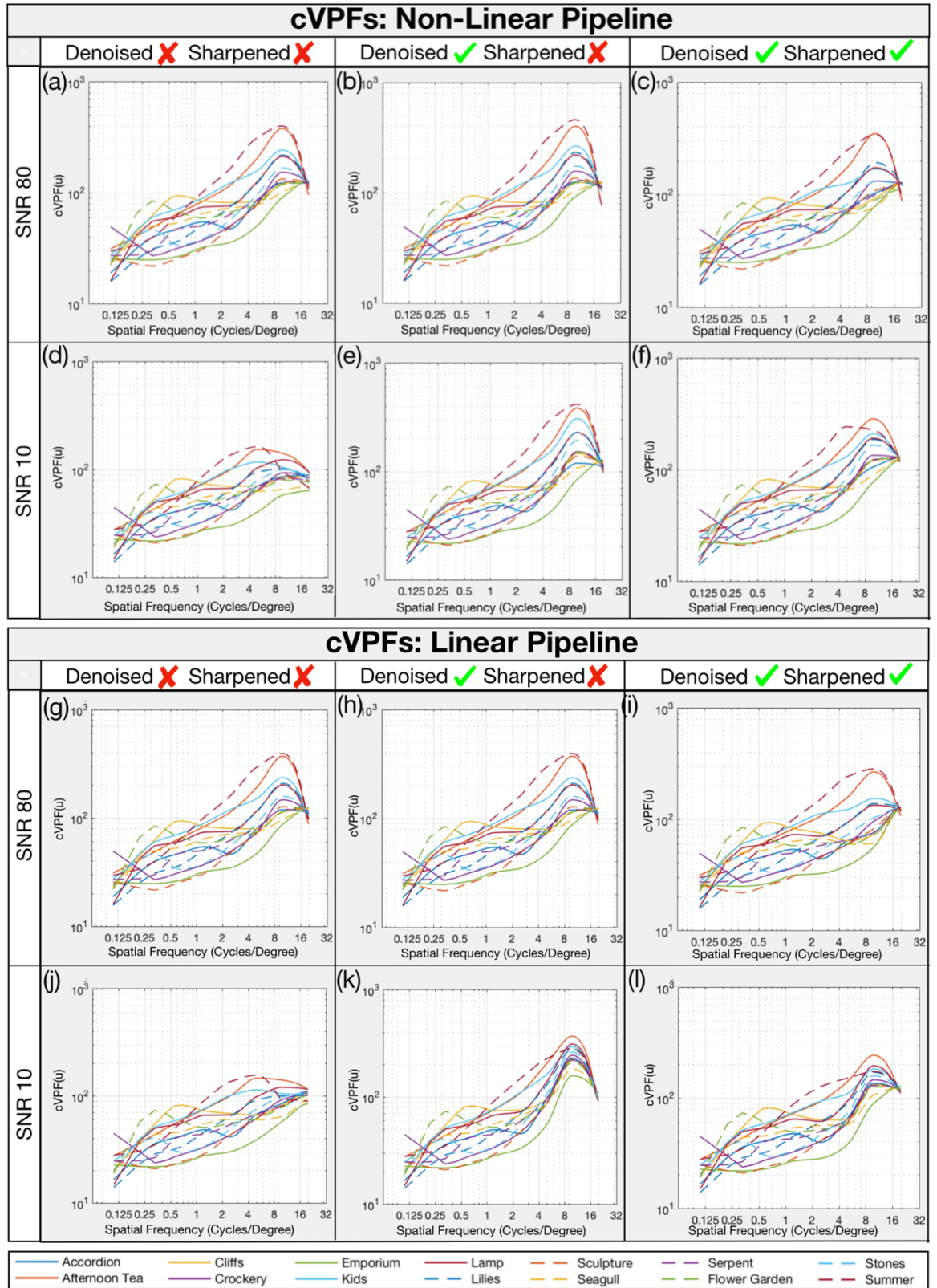


Figure I4 Contextual Visual Perception Functions (cVPF) [8] that were used as input parameters to the image quality metric (IQM) variants in Chapter 6. cVPFs for each input scene image are presented after relevant stages of linear and non-linear image signal processing (ISP).

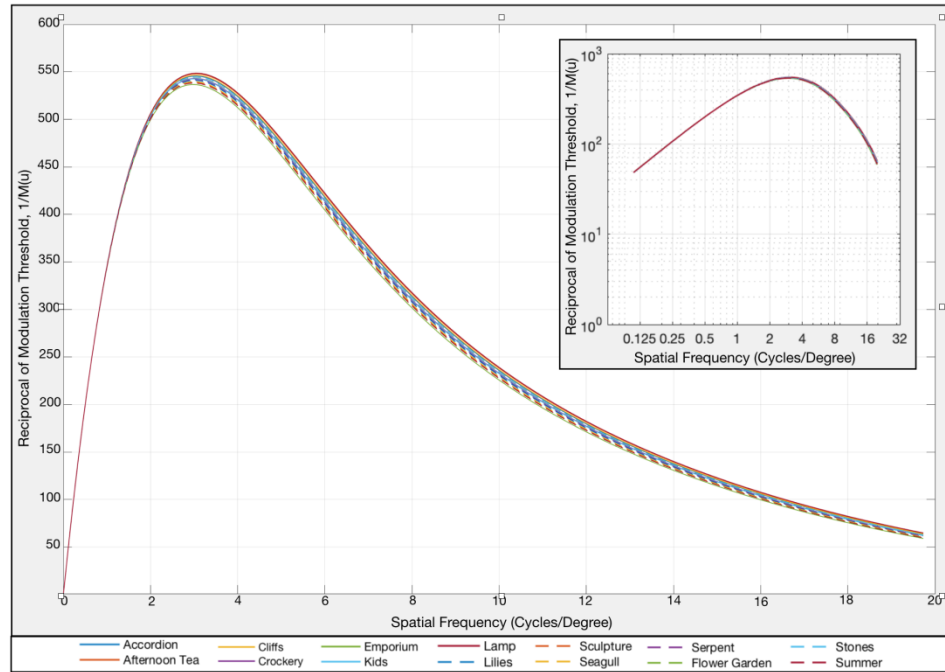


Figure I5 Barten's contrast sensitivity function (CSF) [186] for test images at a signal-to-noise ratio (SNR) of 80, without denoising or sharpening applied. Sensitivities are expressed in a linear-linear space (main plot) and log-log space (inset; \log_{10} y-axis for sensitivity and \log_2 x-axis for spatial frequency). Variation due to image signal processing (ISP) was minimal (not shown).

Reducing the image signal processing (ISP) filter opacity introduced unfiltered noise after denoising and decreased the intensity of sharpening. This affected the various NPS measures shown in Figure I1 in the following ways, which were most significant at lower SNRs:

- 1) A noise floor was introduced to the uniform patch NPS.
- 2) Certain pictorial image SPD-NPS measurements tended toward this noise floor at mid-to-high frequencies for the non-linear pipeline.
- 3) The noise power was higher after denoising compared to the full-opacity pipelines.
- 4) Scene-dependent spread in the non-linear pipeline's SPD-NPS measurements was lower compared to when the ISP filters were tuned full opacity (as shown in Figure 4.6).

The MTF measures were affected in the following ways by reducing the ISP filter opacity (Figure I2). Scene-dependent spread in the SPD-MTFs was lower than at full opacity (Figure 5.4). After denoising and sharpening, levels of bias in the SPD-MTFs were also higher compared to full opacity. This bias was due to increased levels of noise, which was underestimated by all SPD-NPS measures. Using more replicates would mitigate this bias.

Conversely, reducing the ISP opacity lowered bias in the direct dead leaves MTF for the non-linear pipeline (Figure I2(k) and (l)). This was because the uniform patch NPS measured unfiltered noise appropriately but underestimated noise filtered by non-linear content-aware denoising.

Barten's CSF [186] accounted for the luminance of the graphics user interface (GUI) (Figure I5). The latter was calculated from a weighted average of the mean test image luminance and the background luminance of the GUI, as predicted using a Gain Offset Gamma (GOG) model of the standardised RGB (sRGB) colour space (Equation 2.5) [131]. The Contextual CSF (cCSF; Figure I3) [35] and Contextual Visual Perception Function (cVPF; Figure I4) [8] also accounted for the above. Their shape and magnitude were similar to one another, at the different ISP stages. They were more scene-dependent than Barten's CSF since they also accounted for visual masking. Unfiltered noise at SNR 10 impeded detection and discrimination of image signals (Figures I3 and I4, (d) and (j)) but this was mitigated by denoising. Since linear denoising removed virtually all high frequencies from the image at SNR 10, it reduced the effect of masking at these frequencies and raised the cCSF and cVPF.

Appendix J Mean Absolute Errors of IQMs

Tables J1 to J6 benchmark the Mean Absolute Error (MAE) of all metric variants and are discussed in Section 6.2.2. Columns 1 to 3 list the Noise Power Spectrum (NPS), Modulation Transfer Function (MTF) and contrast sensitivity function (CSF) input parameters of each image quality metric (IQM) variant that are summarised in Tables 4.1, 5.1 and 6.1 respectively. Note that the *Combination CSF* refers to the mean of the Contextual CSF (cCSF) [35] (Equation 2.59) and Contextual Visual Perception Function (cVPF) [8] (Equation 2.60). The *Standard CSF* refers to Johnson and Fairchild’s [188] luminance CSF from the IEEE P1858 Camera Phone Image Quality (CPIQ) standard [22] (Equation 2.54). MAE scores were plotted in Columns 4-11 with respect to the following:

Column 4 – the full observer and scene image dataset (i.e. data for all observers and all scene images).

Columns 5 to 7 – data subsets with respect to different combinations of scene susceptibility and observer sensitivity groups defined in Section 6.2.1.

Columns 8 to 9 – data subsets with respect to different observer experience groups.

Column 10 – the full observer and scene image dataset. Each IQM variant was calculated without the CSF/cCSF/cVPF being normalised as described in Section 6.1.3.

Column 11 – the full observer and scene image dataset. The range of integration over the MTF, NPS and CSF were restricted to $12 < u \leq \infty$ cycles/degree to mitigate the effect of the most biased frequencies of the scene-and-process-dependent MTF (SPD-MTF) measures on the IQMs (discussed in Chapter 5). Thus, the resultant IQMs only account for information between $0 < u \leq 12$ cycles/degree. This decision was justified since the most important spatial frequency range for “seeing objects” is between 0 and 15 cycles/degree [334].

Note that Columns 3 and 10 were omitted when comparing the log Noise Equivalent Quanta (log NEQ) metric variants that did not implement any CSF (i.e. Table J4). Column 10 was also omitted when comparing variants of the CPIQ metric (Table J1) since the CSF is present in both the numerator and denominator of the Acutance measure of the CPIQ metric (Equation 3.1) and is effectively normalised by default.

CPIQ Metric: Non-Linear Pipeline									
Noise Image Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (MTF & NPS Domain Restricted to 0 < u < 12 cycles/degree)
Pictorial Image Replicates	Pictorial Image SPD-MTF	Johnson et al.	5.92	6.48	6.50	5.94	6.24	5.01	12.24
Pictorial Image Replicates	Pictorial Image SPD-MTF	cCSF	5.12	5.64	5.61	5.15	5.44	4.24	10.32
Pictorial Image Replicates	Pictorial Image SPD-MTF	cVPF	6.54	7.13	7.17	6.55	6.85	5.61	13.73
Pictorial Image Replicates	Pictorial Image SPD-MTF	Barten	1.82	1.87	1.73	1.87	1.84	2.18	1.80
Pictorial Image Replicates	Pictorial Image SPD-MTF	Johnson et al.	2.55	2.72	2.63	2.58	2.82	2.09	4.26
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Combination	6.22	6.85	6.85	6.24	6.55	5.28	12.16
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	cCSF	5.40	5.96	5.93	5.44	5.70	4.54	10.31
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	cVPF	6.87	7.51	7.51	6.87	7.20	5.86	13.61
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Barten	1.84	1.82	1.69	1.89	1.89	2.13	1.89
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Johnson et al.	2.70	2.85	2.77	2.75	2.96	2.21	4.30
Pictorial Image Replicates	Dead Leaves SPD-MTF	Combination	8.67	9.26	9.24	8.68	9.00	7.66	12.46
Pictorial Image Replicates	Dead Leaves SPD-MTF	cCSF	7.84	8.43	8.37	7.84	8.16	6.88	11.07
Pictorial Image Replicates	Dead Leaves SPD-MTF	cVPF	9.30	9.91	9.89	9.31	9.63	8.27	13.59
Pictorial Image Replicates	Dead Leaves SPD-MTF	Barten	2.26	2.47	2.43	2.31	2.44	2.09	2.65
Pictorial Image Replicates	Dead Leaves SPD-MTF	Johnson et al.	4.48	4.71	4.66	4.47	4.78	3.71	5.78
Dead Leaves Replicates	Dead Leaves SPD-MTF	Combination	8.91	9.37	9.33	8.91	9.23	7.82	12.62
Dead Leaves Replicates	Dead Leaves SPD-MTF	cCSF	8.04	8.51	9.57	8.03	8.36	7.00	11.23
Dead Leaves Replicates	Dead Leaves SPD-MTF	cVPF	9.44	9.96	9.95	9.44	9.77	8.35	13.68
Dead Leaves Replicates	Dead Leaves SPD-MTF	Barten	2.29	2.46	2.41	2.35	2.48	2.07	2.69
Dead Leaves Replicates	Dead Leaves SPD-MTF	Johnson et al.	4.55	4.70	4.64	4.54	4.86	3.70	5.86
Dead Leaves Replicates	Direct Dead Leaves MTF	Combination	9.30	9.81	9.79	9.29	9.62	8.21	12.51
Dead Leaves Replicates	Direct Dead Leaves MTF	cCSF	8.41	8.92	8.86	8.40	8.72	7.34	11.11
Dead Leaves Replicates	Direct Dead Leaves MTF	cVPF	9.88	10.43	10.45	9.88	10.21	8.79	13.58
Dead Leaves Replicates	Direct Dead Leaves MTF	Barten	2.20	2.37	2.30	2.26	2.39	2.00	2.44
Dead Leaves Replicates	Direct Dead Leaves MTF	Johnson et al.	4.43	4.59	4.52	4.41	4.74	3.56	5.48
Uniform Patch	Direct Dead Leaves MTF	Combination	8.83	9.41	9.44	8.81	9.12	7.88	12.13
Uniform Patch	Direct Dead Leaves MTF	cCSF	7.94	8.54	8.56	7.92	8.22	7.02	10.61
Uniform Patch	Direct Dead Leaves MTF	cVPF	9.49	10.07	10.14	9.46	9.79	8.52	13.36
Uniform Patch	Direct Dead Leaves MTF	Barten	1.99	2.15	2.23	2.05	2.08	2.12	2.18
Uniform Patch	Direct Dead Leaves MTF	Johnson et al.	3.96	4.21	4.27	3.96	4.22	3.30	4.93

Least Scene-Dependent

Most Scene-Dependent

CPIQ Metric: Linear Pipeline									
Noise Image Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (MTF & NPS Domain Restricted to 0 < u < 12 cycles/degree)
Pictorial Image Replicates	Pictorial Image SPD-MTF	Johnson et al.	6.90	7.71	7.73	6.77	7.21	5.91	11.36
Pictorial Image Replicates	Pictorial Image SPD-MTF	cCSF	6.21	6.97	6.95	6.09	6.52	5.25	9.86
Pictorial Image Replicates	Pictorial Image SPD-MTF	cVPF	7.41	8.26	8.30	7.29	7.73	6.42	12.61
Pictorial Image Replicates	Pictorial Image SPD-MTF	Barten	1.64	1.71	1.59	1.78	1.78	1.71	1.70
Pictorial Image Replicates	Pictorial Image SPD-MTF	Johnson et al.	2.97	3.23	3.08	2.89	3.27	2.29	4.17
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Combination	7.19	7.75	7.57	7.04	7.51	6.09	11.52
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	cCSF	7.19	7.01	6.82	6.31	6.79	5.39	10.01
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	cVPF	6.46	8.31	8.12	7.59	8.07	6.63	12.79
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Barten	1.62	1.73	1.59	1.76	1.79	1.65	1.73
Pictorial Image Replicates	Mean Pictorial Image SPD-MTF	Johnson et al.	3.11	3.28	3.03	3.02	3.43	2.32	4.27
Pictorial Image Replicates	Dead Leaves SPD-MTF	Combination	7.17	7.74	7.57	7.02	7.50	6.09	11.39
Pictorial Image Replicates	Dead Leaves SPD-MTF	cCSF	6.43	6.98	8.18	6.28	6.75	5.37	9.86
Pictorial Image Replicates	Dead Leaves SPD-MTF	cVPF	7.76	8.34	8.18	7.61	8.09	6.65	12.65
Pictorial Image Replicates	Dead Leaves SPD-MTF	Barten	1.59	1.68	1.55	1.73	1.75	1.64	1.70
Pictorial Image Replicates	Dead Leaves SPD-MTF	Johnson et al.	3.05	3.21	2.97	2.97	3.36	2.29	4.17
Dead Leaves Replicates	Dead Leaves SPD-MTF	Combination	7.17	7.72	7.57	7.01	7.49	6.08	11.38
Dead Leaves Replicates	Dead Leaves SPD-MTF	cCSF	6.43	6.96	6.77	6.28	6.76	5.37	9.88
Dead Leaves Replicates	Dead Leaves SPD-MTF	cVPF	7.67	8.29	8.17	7.52	8.00	6.59	12.58
Dead Leaves Replicates	Dead Leaves SPD-MTF	Barten	1.57	1.66	1.53	1.71	1.72	1.65	1.65
Dead Leaves Replicates	Dead Leaves SPD-MTF	Johnson et al.	2.97	3.13	2.89	2.89	3.28	2.22	4.11
Dead Leaves Replicates	Direct Dead Leaves MTF	Combination	8.19	8.84	8.69	8.04	8.52	7.05	11.69
Dead Leaves Replicates	Direct Dead Leaves MTF	cCSF	7.30	7.89	7.72	7.15	7.62	6.18	10.23
Dead Leaves Replicates	Direct Dead Leaves MTF	cVPF	8.82	9.52	9.40	8.66	9.14	7.69	12.89
Dead Leaves Replicates	Direct Dead Leaves MTF	Barten	1.64	1.77	1.62	1.76	1.83	1.64	1.73
Dead Leaves Replicates	Direct Dead Leaves MTF	Johnson et al.	3.35	3.53	3.29	3.24	3.68	2.48	4.35
Uniform Patch	Direct Dead Leaves MTF	Combination	8.28	8.92	8.77	8.13	8.61	7.14	11.75
Uniform Patch	Direct Dead Leaves MTF	cCSF	7.40	7.99	7.82	7.25	7.73	6.28	10.30
Uniform Patch	Direct Dead Leaves MTF	cVPF	8.90	9.59	9.48	8.74	9.22	7.77	12.93
Uniform Patch	Direct Dead Leaves MTF	Barten	1.72	1.86	1.70	1.81	1.92	1.64	1.83
Uniform Patch	Direct Dead Leaves MTF	Johnson et al.	3.48	3.66	3.42	3.36	3.81	2.57	4.46

Least Scene-Dependent

Most Scene-Dependent

← Lowest MAE of CPIQ Metric Variants Across Both Pipelines	Maximum	9.88	10.43	10.45	9.88	10.21	8.79	13.73
	Minimum	1.57	1.66	1.53	1.71	1.72	1.64	1.65
	Mean	5.65	6.07	6.03	5.62	5.94	4.89	8.44
	Standard Dev.	2.68	2.89	2.95	2.64	2.74	2.33	4.36
← Highest MAE of CPIQ Metric Variants Across Both Pipelines	Relative St.Dev.	47.48%	47.62%	48.96%	46.94%	46.18%	47.67%	51.72%

Table J1 Mean Absolute Error (MAE) between the IEEE P1858 Camera Phone Image Quality (CPIQ) metric variant scores for test images generated by the non-linear (top) and linear pipeline (bottom), and the perceived quality rating of these images.

Visual Log NEQ: Non-Linear Pipeline										
NPS Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (CSF Parameter was not Normalized)	Observers: All Scenes: All (MTF & NPS Domain Restricted to $0 < u < 12$ cycles/degree)
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.01	1.68	1.48	2.11	1.91	2.66	2.12	1.99
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.04	1.70	1.50	2.15	1.94	2.69	2.16	2.03
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	1.87	1.56	1.37	1.99	1.77	2.52	1.87	1.86
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	1.98	1.65	1.44	2.09	1.88	2.64	2.09	1.97
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.02	1.67	1.47	2.12	1.91	2.68	2.13	2.01
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	1.86	1.54	1.34	1.98	1.77	2.51	1.86	1.85
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	1.74	1.41	1.22	1.88	1.65	2.41	1.86	1.72
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	1.77	1.44	1.25	1.91	1.68	2.45	1.91	1.75
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	1.71	1.41	1.21	1.84	1.63	2.37	1.71	1.71
Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cCSF	1.75	1.42	1.22	1.88	1.67	2.40	1.89	1.74
Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cVPF	1.78	1.45	1.25	1.91	1.70	2.42	1.94	1.77
Pictorial Image SPD-NPS	Direct Dead Leaves MTF	Barten	1.73	1.42	1.22	1.85	1.64	2.39	1.73	1.73
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	1.95	1.67	1.48	2.06	1.85	2.58	2.07	1.93
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	1.98	1.69	1.50	2.09	1.88	2.61	2.10	1.97
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	1.83	1.55	1.37	1.95	1.73	2.47	1.83	1.82
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	1.92	1.63	1.44	2.04	1.83	2.56	2.03	1.92
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	1.95	1.65	1.46	2.06	1.85	2.59	2.07	1.95
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	1.82	1.54	1.35	1.94	1.73	2.46	1.82	1.81
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	1.69	1.40	1.22	1.83	1.60	2.34	1.82	1.68
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	1.72	1.43	1.25	1.86	1.62	2.37	1.85	1.70
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	1.68	1.40	1.22	1.81	1.60	2.33	1.68	1.68
Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cCSF	1.70	1.42	1.23	1.83	1.63	2.33	1.84	1.70
Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cVPF	1.72	1.44	1.26	1.85	1.65	2.35	1.89	1.72
Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	Barten	1.70	1.42	1.23	1.82	1.61	2.34	1.70	1.70
Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.11	1.83	1.64	2.21	2.01	2.72	2.23	2.07
Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.14	1.85	1.67	2.25	2.04	2.76	2.27	2.12
Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	Barten	1.91	1.63	1.44	2.02	1.81	2.53	1.91	1.89
Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.08	1.79	1.60	2.19	1.98	2.70	2.20	2.06
Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.12	1.82	1.63	2.22	2.02	2.75	2.24	2.11
Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	1.90	1.61	1.42	2.01	1.80	2.52	1.90	1.89
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cCSF	1.81	1.52	1.33	1.93	1.71	2.44	1.91	1.79
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cVPF	1.84	1.55	1.36	1.96	1.74	2.48	1.95	1.82
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Barten	1.74	1.46	1.27	1.87	1.66	2.37	1.74	1.74
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cCSF	1.83	1.55	1.36	1.94	1.75	2.45	1.96	1.82
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cVPF	1.87	1.58	1.39	1.97	1.78	2.48	2.01	1.85
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Barten	1.76	1.48	1.29	1.88	1.67	2.39	1.76	1.75
Uniform Patch NPS	Pictorial Image SPD-MTF	cCSF	2.47	2.20	2.07	2.55	2.35	3.18	2.57	2.49
Uniform Patch NPS	Pictorial Image SPD-MTF	cVPF	2.48	2.19	2.06	2.56	2.36	3.18	2.57	2.51
Uniform Patch NPS	Pictorial Image SPD-MTF	Barten	2.41	2.15	2.04	2.48	2.29	3.13	2.41	2.41
Uniform Patch NPS	Mean Pictorial Image SPD-MTF	cCSF	2.46	2.17	2.04	2.53	2.34	3.17	2.54	2.48
Uniform Patch NPS	Mean Pictorial Image SPD-MTF	cVPF	2.46	2.17	2.03	2.54	2.34	3.17	2.55	2.49
Uniform Patch NPS	Mean Pictorial Image SPD-MTF	Barten	2.41	2.14	2.02	2.48	2.29	3.03	2.41	2.41
Uniform Patch NPS	Dead Leaves SPD-MTF	cCSF	2.26	1.98	1.87	2.35	2.14	3.00	2.33	2.28
Uniform Patch NPS	Dead Leaves SPD-MTF	cVPF	2.25	1.97	1.86	2.34	2.12	2.99	2.32	2.28
Uniform Patch NPS	Dead Leaves SPD-MTF	Barten	2.29	2.03	1.94	2.36	2.17	3.03	2.29	2.29
Uniform Patch NPS	Direct Dead Leaves MTF	cCSF	2.30	2.02	1.92	2.38	2.19	3.02	2.39	2.32
Uniform Patch NPS	Direct Dead Leaves MTF	cVPF	2.29	2.02	1.90	2.38	2.18	3.01	2.39	2.32
Uniform Patch NPS	Direct Dead Leaves MTF	Barten	2.32	2.06	1.96	2.38	2.20	3.05	2.32	2.32

Most Scene-Dependent

Least Scene-Dependent

← Lowest MAE of All Visual Log NEQ Variants Across Both Pipelines

← Highest MAE of All Visual Log NEQ Variants Across Both Pipelines

Maximum	2.48	2.20	2.07	2.56	2.36	3.18	2.57	2.51
Minimum	1.68	1.40	1.21	1.81	1.60	2.33	1.68	1.68
Mean	1.99	1.70	1.52	2.10	1.89	2.65	2.06	1.98
Standard Dev.	0.25	0.26	0.29	0.23	0.24	0.28	0.26	0.26
Relative St.Dev.	12.63%	15.43%	19.10%	11.14%	12.78%	10.60%	12.49%	13.27%

Table J2 Mean Absolute Error (MAE) between the Visual log NEQ metric variant scores for test images generated by the non-linear pipeline, and the perceived quality rating of these images. NB. the minimum, maximum, mean, standard deviation and relative standard deviation (bottom) are calculated with respect to Table J2 only, but the colour coding is with respect to both pipelines, i.e. Tables J2 and J3.

Visual Log NEQ: Linear Pipeline													
			NPS Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (CSF Parameter was not Normalized)	Observers: All Scenes: All (MTF & NPS Domain Restricted to 0 < u < 12 cycles/degree)
Most Scene-Dependent		Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.34	2.09	2.05	2.64	2.26	2.94	2.46	2.30	
		Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.38	2.12	2.08	2.68	2.30	2.99	2.51	2.34	
		Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.30	2.08	2.04	2.60	2.22	2.92	2.30	2.29	
		Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.30	2.09	2.05	2.61	2.22	2.91	2.42	2.28	
		Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.34	2.13	2.09	2.64	2.25	2.95	2.46	2.32	
		Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.29	2.08	2.04	2.60	2.21	2.90	2.29	2.28	
		Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.33	2.12	2.09	2.64	2.24	2.95	2.44	2.29	
		Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.38	2.17	2.14	2.69	2.29	3.01	2.49	2.33	
		Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.29	2.08	2.04	2.60	2.21	2.91	2.29	2.28	
		Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cCSF	2.26	2.05	2.01	2.56	2.19	2.85	2.38	2.26	
Least Scene-Dependent		Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cVPF	2.30	2.09	2.05	2.59	2.22	2.87	2.43	2.29	
		Pictorial Image SPD-NPS	Direct Dead Leaves MTF	Barten	2.27	2.07	2.03	2.58	2.20	2.89	2.27	2.27	
		Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.33	2.08	2.03	2.63	2.25	2.93	2.45	2.29	
		Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.37	2.11	2.07	2.68	2.29	2.97	2.50	2.34	
		Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.28	2.06	2.02	2.59	2.21	2.90	2.28	2.28	
		Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.29	2.08	2.04	2.60	2.22	2.89	2.41	2.27	
		Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.33	2.12	2.08	2.63	2.25	2.93	2.45	2.32	
		Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.27	2.06	2.03	2.59	2.20	2.88	2.27	2.27	
		Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.32	2.11	2.07	2.63	2.24	2.93	2.43	2.28	
		Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.37	2.16	2.13	2.68	2.28	2.99	2.48	2.32	
		Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.28	2.06	2.03	2.59	2.20	2.89	2.28	2.27	
		Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cCSF	2.25	2.04	2.00	2.55	2.18	2.83	2.37	2.25	
		Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	cVPF	2.29	2.08	2.05	2.58	2.22	2.85	2.42	2.29	
		Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	Barten	2.26	2.05	2.02	2.57	2.19	2.87	2.26	2.26	
		Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.33	2.08	2.03	2.63	2.25	2.92	2.44	2.28	
		Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.37	2.11	2.07	2.67	2.29	2.96	2.50	2.33	
		Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	Barten	2.28	2.06	2.02	2.59	2.21	2.89	2.28	2.27	
		Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.29	2.08	2.04	2.60	2.21	2.89	2.40	2.27	
		Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.32	2.11	2.08	2.63	2.24	2.92	2.44	2.31	
		Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.27	2.06	2.02	2.58	2.20	2.88	2.27	2.27	
		Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.32	2.10	2.07	2.62	2.23	2.93	2.42	2.27	
		Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.36	2.15	2.12	2.67	2.28	2.98	2.48	2.32	
		Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Barten	2.27	2.06	2.02	2.58	2.20	2.88	2.27	2.27	
		Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cCSF	2.25	2.04	2.00	2.55	2.18	2.82	2.37	2.24	
		Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cVPF	2.28	2.08	2.04	2.58	2.21	2.85	2.42	2.28	
		Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Barten	2.26	2.05	2.01	2.57	2.19	2.87	2.26	2.26	
		Uniform Patch NPS	Pictorial Image SPD-MTF	cCSF	2.32	2.08	2.03	2.62	2.24	2.91	2.44	2.28	
		Uniform Patch NPS	Pictorial Image SPD-MTF	cVPF	2.37	2.11	2.07	2.67	2.29	2.96	2.50	2.33	
		Uniform Patch NPS	Pictorial Image SPD-MTF	Barten	2.28	2.06	2.02	2.59	2.20	2.88	2.28	2.27	
		Uniform Patch NPS	Mean Pictorial Image SPD-MTF	cCSF	2.29	2.08	2.04	2.59	2.21	2.88	2.40	2.27	
		Uniform Patch NPS	Mean Pictorial Image SPD-MTF	cVPF	2.32	2.11	2.08	2.63	2.24	2.92	2.44	2.31	
		Uniform Patch NPS	Mean Pictorial Image SPD-MTF	Barten	2.27	2.06	2.02	2.58	2.20	2.87	2.27	2.26	
		Uniform Patch NPS	Dead Leaves SPD-MTF	cCSF	2.31	2.10	2.07	2.62	2.23	2.92	2.42	2.27	
		Uniform Patch NPS	Dead Leaves SPD-MTF	cVPF	2.36	2.15	2.12	2.67	2.27	2.98	2.48	2.32	
		Uniform Patch NPS	Dead Leaves SPD-MTF	Barten	2.27	2.06	2.02	2.58	2.20	2.87	2.27	2.26	
		Uniform Patch NPS	Direct Dead Leaves MTF	cCSF	2.25	2.04	2.00	2.54	2.18	2.82	2.37	2.24	
		Uniform Patch NPS	Direct Dead Leaves MTF	cVPF	2.28	2.08	2.04	2.58	2.21	2.84	2.42	2.28	
		Uniform Patch NPS	Direct Dead Leaves MTF	Barten	2.26	2.05	2.01	2.57	2.19	2.86	2.26	2.26	

Table J3 Mean Absolute Error (MAE) between the Visual log NEQ metric variant scores for test images generated by the linear pipeline, and the perceived quality rating of these images. NB. the minimum, maximum, mean, standard deviation and relative standard deviation (bottom) are calculated with respect to Table J3 only, but the colour coding is with respect to both pipelines, i.e. Tables J2 and J3.

Log NEQ: Non-Linear Pipeline								
NPS Employed	MTF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (MTF & NPS Domain Restricted to $0 < u < 12$ cycles/degree)
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	2.30	2.20	2.01	2.40	2.36	2.42	2.32
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	2.28	2.17	1.96	2.38	2.36	2.37	2.34
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	2.16	2.15	2.00	2.27	2.27	2.13	2.29
Pictorial Image SPD-NPS	Direct Dead Leaves MTF	2.24	2.26	2.10	2.35	2.36	2.18	2.26
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	2.24	2.20	2.02	2.36	2.31	2.35	2.27
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	2.22	2.18	1.99	2.34	2.30	2.30	2.29
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	2.20	2.21	2.08	2.33	2.33	2.10	2.33
Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	2.26	2.31	2.17	2.39	2.41	2.14	2.30
Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	2.41	2.35	2.12	2.49	2.44	2.56	2.37
Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	2.38	2.32	2.11	2.46	2.43	2.50	2.38
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	2.22	2.21	2.05	2.34	2.34	2.19	2.35
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	2.27	2.29	2.12	2.40	2.40	2.24	2.31
Uniform Patch NPS	Pictorial Image SPD-MTF	3.00	2.96	2.88	3.03	2.97	3.44	3.07
Uniform Patch NPS	Mean Pictorial Image SPD-MTF	3.01	2.96	2.86	3.03	2.99	3.41	3.07
Uniform Patch NPS	Dead Leaves SPD-MTF	2.75	2.73	2.65	2.75	2.72	3.16	2.84
Uniform Patch NPS	Direct Dead Leaves MTF	2.87	2.86	2.79	2.86	2.85	3.25	2.91

Log NEQ: Linear Pipeline								
NPS Employed	MTF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (MTF & NPS Domain Restricted to $0 < u < 12$ cycles/degree)
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	2.12	2.04	1.95	2.33	2.16	2.27	2.06
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	2.06	2.05	1.95	2.27	2.12	2.16	2.05
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	2.07	2.05	1.95	2.29	2.12	2.20	2.04
Pictorial Image SPD-NPS	Direct Dead Leaves MTF	2.04	2.05	1.93	2.24	2.12	2.12	2.04
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	2.13	2.05	1.96	2.33	2.17	2.26	2.07
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	2.06	2.06	1.96	2.27	2.13	2.16	2.06
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	2.07	2.05	1.96	2.29	2.13	2.19	2.06
Mean Pictorial Image SPD-NPS	Direct Dead Leaves MTF	2.05	2.07	1.95	2.24	2.14	2.12	2.05
Dead Leaves SPD-NPS	Pictorial Image SPD-MTF	2.12	2.05	1.94	2.33	2.17	2.26	2.07
Dead Leaves SPD-NPS	Mean Pictorial Image SPD-MTF	2.06	2.06	1.95	2.26	2.13	2.16	2.06
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	2.07	2.05	1.96	2.28	2.13	2.19	2.05
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	2.05	2.06	1.94	2.24	2.13	2.12	2.05
Uniform Patch NPS	Pictorial Image SPD-MTF	2.13	2.06	1.97	2.33	2.18	2.26	2.08
Uniform Patch NPS	Mean Pictorial Image SPD-MTF	2.07	2.07	1.96	2.26	2.14	2.16	2.07
Uniform Patch NPS	Dead Leaves SPD-MTF	2.07	2.06	1.96	2.28	2.14	2.18	2.06
Uniform Patch NPS	Direct Dead Leaves MTF	2.06	2.08	1.95	2.24	2.15	2.11	2.06

← Lowest MAE of All Log NEQ Variants Across Both Pipelines		Maximum	3.01	2.96	2.88	3.03	2.99	3.44	3.07
		Minimum	2.04	2.04	1.93	2.24	2.12	2.10	2.04
		Mean	2.25	2.23	2.10	2.40	2.32	2.36	2.27
		Standard Dev.	0.27	0.26	0.27	0.21	0.24	0.38	0.29
← Highest MAE of All Log NEQ Variants Across Both Pipelines		Relative St.Dev.	12.02%	11.87%	12.96%	8.80%	10.50%	15.94%	12.97%

Table J4 Mean Absolute Error (MAE) between the log Noise Equivalent Quanta (log NEQ) metric variant scores for test images generated by the non-linear (top) and linear pipeline (bottom), and the perceived quality rating of these images.

PIC: Non-Linear Pipeline										
NPS Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (CSF Parameter was not Normalized)	Observers: All Scenes: All (MTF & NPS Domain Restricted to 0 < u < 12 cycles/degree)
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Combination	2.63	2.36	2.09	2.64	2.75	2.50	2.63	2.25
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.63	2.36	2.09	2.64	2.75	2.50	2.63	2.25
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.63	2.36	2.09	2.64	2.75	2.50	2.63	2.25
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.63	2.36	2.09	2.64	2.75	2.50	2.63	2.25
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Combination	2.59	2.31	2.02	2.61	2.72	2.51	2.59	2.23
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.59	2.31	2.02	2.61	2.72	2.51	2.59	2.23
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.59	2.31	2.02	2.61	2.72	2.51	2.59	2.23
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.59	2.31	2.02	2.61	2.72	2.51	2.59	2.23
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Combination	2.50	2.18	1.88	2.55	2.70	2.21	2.50	2.20
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.50	2.18	1.88	2.55	2.70	2.21	2.50	2.20
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.50	2.18	1.88	2.55	2.70	2.21	2.50	2.20
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.50	2.18	1.88	2.55	2.70	2.21	2.50	2.20
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Combination	2.76	2.38	2.10	2.80	2.91	2.58	2.76	2.36
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.76	2.38	2.10	2.80	2.91	2.58	2.76	2.36
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.76	2.38	2.10	2.80	2.91	2.58	2.76	2.36
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.76	2.38	2.10	2.80	2.91	2.58	2.76	2.36
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Combination	2.75	2.35	2.05	2.78	2.88	2.57	2.75	2.34
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.75	2.35	2.05	2.78	2.88	2.57	2.75	2.34
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.75	2.35	2.05	2.78	2.88	2.57	2.75	2.34
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.75	2.35	2.05	2.78	2.88	2.57	2.75	2.34
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Combination	2.69	2.27	1.96	2.73	2.88	2.31	2.69	2.35
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.69	2.27	1.96	2.73	2.88	2.31	2.69	2.35
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.69	2.27	1.96	2.73	2.88	2.31	2.69	2.35
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.69	2.27	1.96	2.73	2.88	2.31	2.69	2.35
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Combination	2.66	2.24	1.90	2.71	2.84	2.40	2.66	2.33
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.66	2.24	1.90	2.71	2.84	2.40	2.66	2.33
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.66	2.24	1.90	2.71	2.84	2.40	2.66	2.33
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Barten	2.66	2.24	1.90	2.71	2.84	2.40	2.66	2.33
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Combination	2.72	2.36	2.08	2.76	2.87	2.50	2.72	2.32
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cCSF	2.72	2.36	2.08	2.76	2.87	2.50	2.72	2.32
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cVPF	2.72	2.36	2.08	2.76	2.87	2.50	2.72	2.32
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Barten	2.72	2.36	2.08	2.76	2.87	2.50	2.72	2.32
Uniform Patch NPS	Direct Dead Leaves MTF	Combination	3.15	3.15	3.11	3.14	3.21	3.27	3.15	2.81
Uniform Patch NPS	Direct Dead Leaves MTF	cCSF	3.15	3.15	3.11	3.14	3.21	3.27	3.15	2.81
Uniform Patch NPS	Direct Dead Leaves MTF	cVPF	3.15	3.15	3.11	3.14	3.21	3.27	3.15	2.81
Uniform Patch NPS	Direct Dead Leaves MTF	Barten	3.15	3.15	3.11	3.14	3.21	3.27	3.15	2.81

PIC: Linear Pipeline										
NPS Employed	MTF Employed	CSF Employed	Observers: All Scenes: All	Observers: All Scenes: Medium and Low Susceptibility	Observers: All Scenes: Medium Susceptibility	Observers: Medium Sensitivity Scenes: All	Observers: Inexperienced Scenes: All	Observers: Experienced Scenes: All	Observers: All Scenes: All (CSF Parameter was not Normalized)	Observers: All Scenes: All (MTF & NPS Domain Restricted to 0 < u < 12 cycles/degree)
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Combination	2.47	2.12	1.90	2.60	2.60	2.31	2.47	2.28
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.47	2.12	1.90	2.60	2.60	2.31	2.47	2.28
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.47	2.12	1.90	2.60	2.60	2.31	2.47	2.28
Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.47	2.12	1.90	2.60	2.60	2.31	2.47	2.28
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Combination	2.51	2.14	1.91	2.65	2.65	2.34	2.51	2.30
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.51	2.14	1.91	2.65	2.65	2.34	2.51	2.30
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.51	2.14	1.91	2.65	2.65	2.34	2.51	2.30
Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.51	2.14	1.91	2.65	2.65	2.34	2.51	2.30
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Combination	2.47	2.12	1.90	2.62	2.60	2.34	2.47	2.29
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.47	2.12	1.90	2.62	2.60	2.34	2.47	2.29
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.47	2.12	1.90	2.62	2.60	2.34	2.47	2.29
Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.47	2.12	1.90	2.62	2.60	2.34	2.47	2.29
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Combination	2.48	2.13	1.90	2.62	2.63	2.31	2.48	2.29
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cCSF	2.48	2.13	1.90	2.62	2.63	2.31	2.48	2.29
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	cVPF	2.48	2.13	1.90	2.62	2.63	2.31	2.48	2.29
Mean Pictorial Image SPD-NPS	Pictorial Image SPD-MTF	Barten	2.48	2.13	1.90	2.62	2.63	2.31	2.48	2.29
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Combination	2.53	2.15	1.91	2.66	2.67	2.34	2.53	2.31
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cCSF	2.53	2.15	1.91	2.66	2.67	2.34	2.53	2.31
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	cVPF	2.53	2.15	1.91	2.66	2.67	2.34	2.53	2.31
Mean Pictorial Image SPD-NPS	Mean Pictorial Image SPD-MTF	Barten	2.53	2.15	1.91	2.66	2.67	2.34	2.53	2.31
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Combination	2.49	2.12	1.90	2.63	2.62	2.34	2.49	2.30
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.49	2.12	1.90	2.63	2.62	2.34	2.49	2.30
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.49	2.12	1.90	2.63	2.62	2.34	2.49	2.30
Mean Pictorial Image SPD-NPS	Dead Leaves SPD-MTF	Barten	2.49	2.12	1.90	2.63	2.62	2.34	2.49	2.30
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Combination	2.48	2.12	1.90	2.63	2.62	2.33	2.48	2.29
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cCSF	2.48	2.12	1.90	2.63	2.62	2.33	2.48	2.29
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	cVPF	2.48	2.12	1.90	2.63	2.62	2.33	2.48	2.29
Dead Leaves SPD-NPS	Dead Leaves SPD-MTF	Barten	2.48	2.12	1.90	2.63	2.62	2.33	2.48	2.29
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Combination	2.51	2.15	1.90	2.63	2.66	2.30	2.51	2.29
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cCSF	2.51	2.15	1.90	2.63	2.66	2.30	2.51	2.29
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	cVPF	2.51	2.15	1.90	2.63	2.66	2.30	2.51	2.29
Dead Leaves SPD-NPS	Direct Dead Leaves MTF	Barten	2.51	2.15	1.90	2.63	2.66	2.30	2.51	2.29
Uniform Patch NPS	Direct Dead Leaves MTF	Combination	2.52	2.16	1.90	2.65	2.68	2.31	2.52	2.31
Uniform Patch NPS	Direct Dead Leaves MTF	cCSF	2.52	2.16	1.90	2.65	2.68	2.31	2.52	2.31
Uniform Patch NPS	Direct Dead Leaves MTF	cVPF	2.52	2.16	1.90	2.65	2.68	2.31	2.52	2.31
Uniform Patch NPS	Direct Dead Leaves MTF	Barten	2.52	2.16	1.90	2.65	2.68	2.31	2.52	2.31

← Lowest MAE of All PIC Variants Across Both Pipelines

Highest MAE of All PIC Variants Across Both Pipelines

Maximum	3.15	3.15	3.11	3.14	3.21	3.27	3.15	2.81
Minimum	2.47	2.12	1.88	2.55	2.60	2.21	2.47	2.20
Mean	2.61	2.27	2.02	2.69	2.75	2.43	2.61	2.33
Standard Dev.	0.16	0.23	0.27	0.13	0.15	0.23	0.16	0.12
Relative St.Dev.	6.31%	10.31%	13.60%	4.75%	5.57%	9.32%	6.31%	5.29%

Table J5 Mean Absolute Error (MAE) between the Perceived Information Capacity (PIC) variant scores for test images generated by the non-linear (top) and linear (bottom) pipelines, and the perceived quality rating of these images.

Table J6 Mean Absolute Error (MAE) between the Square Root Integral with Noise (SQRIn) variant scores for test images generated by the non-linear (top) and linear (bottom) pipelines, and the perceived quality rating of these images.

Appendix K Related Work

K.1 List of Publications

E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, J. R. Jarvis, and R. E. Jacobson, “Validation of Modulation Transfer Functions and Noise Power Spectra from Natural Scenes,” *J. Imaging Sci. Technol.* vol. 63, no. 6, (in print), 2020. *

E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, “Scene-and-Process-Dependent Spatial Image Quality Metrics,” *J. Imaging Sci. Technol.*, vol. 63, no. 6, (in print), 2020. *

E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, “Noise Power Spectrum Scene-Dependency in Simulated Image Capture Systems,” accepted for publication in *Proc. IS&T Electronic Imaging, Image Quality and System Performance XVII*, 2020. *

E. W. S. Fry, S. Triantaphillidou, R. E. Jacobson, J. R. Jarvis, and R. B. Jenkin, “Bridging the Gap Between Imaging Performance and Image Quality Measures,” in *Proc. IS&T Electronic Imaging, Image Quality and System Performance XV*, 2018. *

S. Triantaphillidou, J. Smekjal, E. W. S. Fry, and C. H. Hung, “Studies on the effects of megapixel sensor resolution on displayed image quality and relevant metrics,” accepted for publication in *Proc. IS&T Electronic Imaging, Image Quality and System Performance XVII*, 2020.

S. Triantaphillidou, E. W. S. Fry, V. Sanchis-Jurado, and A. Pons, “Image Quality Loss and Compensation for Visually Impaired Observers,” in *Proc. IS&T Electronic Imaging, Image Quality and System Performance XV*, 2018.

E. W. S. Fry, S. Triantaphillidou, J. Jarvis, and G. Gupta, “Image quality optimization, via application of contextual contrast sensitivity and discrimination functions,” in *Proc. SPIE 9396, Image Quality and System Performance XII*, 2015.

* Asterisks indicate publications directly related to the project for which the author of this thesis was the primary author. Selected comments from the reviews of these publications are copied below.

Review of the paper entitled: “*Validation of Modulation Transfer Functions and Noise Power Spectra from Natural Scenes.*”

“This is a sound paper worthy of an audience at the cutting edge of this area of research. I do feel this is work of both rigor and increasing necessity. Congratulations on an exciting piece of work.”

Review of the paper entitled: “*Scene-and-Process-Dependent Spatial Image Quality Metrics.*”

“The authors revised the traditional engineering input parameters and propose two novel image quality metrics. The experiments are detailed and well-desired.”

Review of the paper entitled: “*Noise Power Spectrum Scene-Dependency in Simulated Image Capture Systems.*”

“Very interesting and novel work. Prior art, context and motivation nicely explained. The abstract further develops the concept of scene-and-process-dependent NPS evaluation in image capture systems with content-aware image processing pipelines, building on previous papers. For this, two new noise metrics were developed and tested.”

K.2 Presentations at Conferences and Symposia

E. W. S. Fry (presenter), S. Triantaphillidou, R. B. Jenkin, J. R. Jarvis, and R. E. Jacobson, “Validation of Modulation Transfer Functions and Noise Power Spectra from Natural Scenes,” at *IS&T International Symposium on Electronic Imaging: Image Quality and System Performance XVII*, 2020, San Francisco, USA. *

E. W. S. Fry (presenter), S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, “Scene-and-Process-Dependent Spatial Image Quality Metrics,” at *IS&T International Symposium on Electronic Imaging: Image Quality and System Performance XVII*, 2020, San Francisco, USA. *

E. W. S. Fry (presenter), S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, “Noise Power Spectrum Scene-Dependency in Simulated Image Capture Systems,” at *IS&T International Symposium on Electronic Imaging: Image Quality and System Performance XVII*, 2020, San Francisco, USA. *

S. Triantaphillidou (presenter), E. W. S. Fry (presenter), and O. van Zwanenberg (presenter), “Objective Image Quality Assessment Using Pictures,” at *Transactions Imaging/Art/Science: Image Quality, Content and Aesthetics*, 2019, University of Westminster, Regent Street Campus, London, UK.

E. W. S. Fry (presenter), S. Triantaphillidou, R. E. Jacobson, J. R. Jarvis, and R. B. Jenkin, “Bridging the Gap Between Imaging Performance and Image Quality Measures,” at *IS&T International Symposium on Electronic Imaging: Image Quality and System Performance XV*, 2018, San Francisco, USA.

E. W. S. Fry, “Bridging the Gap Between Imaging Performance and Image Quality Measures,” at Computer Science and Engineering Session at *the University of Westminster 2018 Faculty of Science and Technology Doctoral Conference*, 2018, New Cavendish Street Campus, London, UK.

E. W. S. Fry (presenter), “Visual Image Quality Metrics for Engineers,” at *On Semiconductor (via video link)*, 2017, San Jose, USA.

* Asterisks indicate presentations scheduled for after the date of submission of this thesis, but before the viva.

K.3 Awards

Best Student Paper for the Image Quality and System Performance (IQSP) XV conference, 2018, at the Society for Imaging Science and Technology’s (IS&T) Electronic Imaging Symposium, 2018, San Francisco, USA. Awarded for the paper entitled “*Bridging the Gap Between Imaging Performance and Image Quality Measures*”.

First Place Award for Full-Length Presentation at the Computer Science and Engineering Session at the University of Westminster Faculty of Science and Technology Doctoral Conference, 2018, New Cavendish Street Campus, London, UK. Awarded for the

presentation entitled “*Bridging the Gap Between Imaging Performance and Image Quality Measures*”.

Best Student Paper for the IQSP XII conference, 2015, at the Society of Photo-Optical Instrumentation Engineers (SPIE) & IS&T Electronic Imaging Symposium, San Francisco, USA. Awarded for the paper entitled “*Image quality optimization, via application of contextual contrast sensitivity and discrimination functions*”.

Abbreviations

1D	1-Dimensional
2D	2-Dimensional
3D	3-Dimensional
AC_1C_2	AC_1C_2 colour space
ACF	Autocovariance Function
ADAS	Advanced Driver Assistance System
BKE	Background known exactly
BM3D	Block Matching and 3D Filter
cCSF	Contextual Contrast Sensitivity Function
CFA	Colour filter array
CIE	Commission Internationale de L'Eclairage
CIELAB	CIE 1976 $L^*a^*b^*$ colour space
CIEDE2000	CIE colour difference equation (CIEDE2000)
CIEXYZ	CIE XYZ tristimulus values
CNN	Convolutional neural network
CP-IQM	Computational image quality metric
CPIQ	Camera Phone Image Quality Initiative
CPIQ metric	Metric from the IEEE P1858 Camera Phone Image Quality Standard
CSF	Contrast sensitivity function
CSF/CIEDE2000	CSF/CIEDE2000 image quality metric

cVPF	Contextual Visual Perception Function
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DMOS	Differential Mean Opinion Score
dpi	Dots per inch
DQE	Detective Quantum Efficiency
DSLR	Digital single-lens reflex
DSNU	Dark signal non-uniformity
EOCF	Electro-optic conversion function
EPIC	Effective Perceived Information Capacity
ESF	Edge Spread Function
FBF	Fast Bilateral Filter
FFT	Fast Fourier Transform
FPN	Fixed Pattern Noise
GIF	Guided Image Filter
GOG	Gain Offset Gamma
GUI	Graphics user interface
HVS	Human visual system
iCSF	Isolated Contrast Sensitivity Function
IEEE	Institute of Electrical and Electronics Engineers
IF-IQM	Image fidelity metric
IFC	Information Fidelity Criterion
IQM	Image quality metric

ISO	International Organization for Standardization
ISP	Image signal processing
iVPF	Isolated Visual Perception Function
JPEG	Joint Photographic Experts Group
JPEG 2000	JPEG 2000 image compression standard
JND	Just-noticeable difference
LAM	Linear Amplification Model
LBL	Local band-limited
LCD	Liquid crystal display
Log NEQ	Logarithm of the integral of the Noise Equivalent Quanta
LSF	Line Spread Function
LUT	Look-up table
MAE	Mean Absolute Error
MF-IQM	Multivariate formalism image quality metric
MOS	Mean Opinion Score
MSE	Mean Square Error
MSSIM	Mean Structural Similarity Index
MTF	Modulation Transfer Function
NEQ	Noise Equivalent Quanta
NPS	Noise Power Spectrum
NSS	Natural scene statistics
OCW	Optimal contrast weighting
OECF	Opto-electronic conversion function

OSAP	One Step Alternating Projections
OTF	Optical Transfer Function
PIC	Perceived Information Capacity
PNG	Portable Network Graphics
ppi	Pixels per inch
PRNU	Photoresponse non-uniformity
PSD	Power Spectral Density
PSF	Point Spread Function
PSNR	Peak signal-to-noise ratio
RGB	Red green blue
RMS	Root mean square
RMSE	Root Mean Square Error
ROI	Region of interest
S-CIELAB	Spatial extension to the CIELAB colour space
SFR	Spatial Frequency Response
SKE	Signal known exactly
SNR	Signal-to-noise ratio
SoC	System on Chip
SPD-MTF	Scene-and-process-dependent Modulation Transfer Function
SPD-NEQ	Scene-and-process-dependent Noise Equivalent Quanta
SPD-NPS	Scene-and-process-dependent Noise Power Spectrum
SPD-SNRI	Scene-and-process-dependent Ideal Observer Signal-to-Noise Ratio
SQRI	Square Root Integral

SQRIn	Square Root Integral with Noise
SQS ₂	Standard Quality Scale
sRGB	Standardised RGB colour space
SROCC	Spearman's Rank Order Correlation Coefficient
SSIM	Structural Similarity Index
STV-IQM	Signal transfer visual image quality metric
SVD	Single Value Decomposition
UQI	Universal Quality Index
USM	Unsharp mask
VIF	Visual Information Fidelity
Visual Log NEQ	Visually-weighted logarithm of the integral of the Noise Equivalent Quanta
VNTF	Video noise temporal filtering
WLS	Weighted Least Squares
YCbCr	YCbCr opponent colour space

References

- [1] E. Allen and S. Triantaphillidou, *The Manual of Photography: Photographic and Digital Imaging*, 10th ed. Focal Press, 2011.
- [2] R. B. Jenkin, "On the Application of the Modulation Transfer Function to Discrete Imaging Systems," PhD Thesis, University of Westminster, UK, 2001.
- [3] E. W. Jin *et al.*, "Towards the Development of the IEEE P1858 CPIQ Standard – A validation study," in *Proc. IS&T Electronic Imaging: Image Quality and System Performance XIV*, 2017, pp. 88–94.
- [4] E. Allen, S. Triantaphillidou, and R. E. Jacobson, "Perceptibility and acceptability of JPEG 2000 compressed images of various scene types," in *Proc. SPIE 9016, Image Quality and System Performance XI*, 2014.
- [5] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Winchester MA: Imcotek Press, 2000.
- [6] J. C. Dainty and R. Shaw, *Image Science: Principles, Analysis and Evaluation of Photographic-type Imaging Processes*. London: Academic Press Ltd., 1974.
- [7] R. Branca, S. Triantaphillidou, and P. D. Burns, "Texture MTF from images of natural scenes," in *Proc. IS&T Electronic Imaging: Image Quality and System Performance XIV*, 2017, pp. 113–120.
- [8] S. Triantaphillidou, J. R. Jarvis, and G. Gupta, "Spatial contrast sensitivity and discrimination in pictorial images," in *Proc. SPIE 9016 Image Quality and System Performance XI*, 2014, pp. 901604-1-901604–15.
- [9] A. M. Ford, "Relationships Between Image Quality and Image Compression," PhD Thesis, University of Westminster, UK, 1997.
- [10] R. E. Jacobson, A. M. Ford, and G. G. Attridge, "Evaluation of the effects of compression on the quality of images on a soft display," in *Proc. SPIE. 3016, Human Vision and Electronic Imaging II*, 1997, pp. 114–125.

-
- [11] A. M. Ford, "Determination of Compressed Image Quality," in *Colour Imaging: Vision and Technology*, L. W. Macdonald and M. R. Luo, Eds. Chichester, UK: John Wiley & Sons, 1999, pp. 315–337.
- [12] P. G. J. Barten, "Evaluation of the Effect of Noise on Subjective Image Quality," in *Proc. SPIE 1453, Human Vision, Visual Processing, and Digital Display II*, 1991, pp. 2–15.
- [13] R. E. Jacobson and S. Triantaphillidou, "Metric Approaches to Image Quality," in *Colour Image Science: Exploiting Digital Media*, John Wiley & Sons, 2002, pp. 371–392.
- [14] E. W. S. Fry, S. Triantaphillidou, J. Jarvis, and G. Gupta, "Image quality optimization, via application of contextual contrast sensitivity and discrimination functions," in *Proc. SPIE 9396, Image Quality and System Performance XII*, 2015.
- [15] R. B. Fagard-Jenkin, R. E. Jacobson, and K. Maclellan-Brown, "Determination of the MTF of JPEG Compression Using the ISO 12233 Spatial Frequency Response Plug-in," in *Proc. IS&T PICS Conference: Image Processing, Image Quality, Image Capture Systems*, 2000, pp. 254–258.
- [16] F. Cao, F. Guichard, and H. Hornung, "Dead leaves model for measuring texture quality on a digital camera," in *Proc. SPIE 7537, Digital Photography VI*, 2010, p. 75370E.
- [17] F. Cao, F. Guichard, and H. Hornung, "Measuring Texture Sharpness of A Digital Camera," in *Proc. SPIE 7250, Digital Photography V*, 2009, p. 72500H.
- [18] U. Artmann, "Measurement of Noise Using the Dead Leaves Pattern," in *Proc. IS&T Electronic Imaging: Image Quality and System Performance XV*, 2018, pp. 341-1-341–6.
- [19] J. McElvain, S. P. Campbell, J. Miller, and E. W. Jin, "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," in *Proc. SPIE 7537, Digital Photography VI*, 2010, pp. 75370D-1-75370D–11.
- [20] *ISO 19567-2: 2019: Photography — Digital cameras, Part 2: Texture analysis*

-
- using stochastic pattern*. International Organization for Standardization (ISO), 2019.
- [21] L. Kirk, P. Herzer, U. Artmann, and D. Kunz, "Description of Texture Loss Using the Dead Leaves Target - Current Issues and a New Intrinsic Approach," in *Proc. SPIE 9023, Digit. Photog. X*, 2014.
 - [22] *IEEE Std 1858TM-2016: IEEE Standard for Camera Phone Image Quality*. IEEE Standards Association Board of Governors, 2016.
 - [23] *ISO 12233:2017: Photography — Electronic still picture imaging — resolution and spatial frequency responses*. International Organization for Standardization (ISO), 2017.
 - [24] Imatest, "ISO 12233: 2014 Edge SFR (eSFR) Photographic chart," 2018. [Online]. Available: <http://store.imatest.com/test-charts/iso-cpiq-standard-test-charts/photographic-iso-12233-2014-esfr-chart.html#>. [Accessed: 02-Jan-2019].
 - [25] P. D. Burns, "sfrmat3: SFR evaluation for digital cameras and scanners," 2015. [Online]. Available: http://www.losburns.com/imaging/software/SFRedge/sfrmat3_post/index.html. [Accessed: 04-Dec-2019].
 - [26] Imatest, "Sinusoidal Siemens Star," 2019. [Online]. Available: <http://store.imatest.com/test-charts/iso-cpiq-standard-test-charts/sinusoidal-siemens-star.html>. [Accessed: 17-Mar-2019].
 - [27] Image Engineering, "Index of /software/IE-Resolution," 2016. [Online]. Available: <https://downloads.image-engineering.de/software/IE-Resolution/>. [Accessed: 07-Dec-2019].
 - [28] P. D. . Burns, "Texture MTF: Based on Dead-leaves type of test targets," 2013. [Online]. Available: <http://burnsdigitalimaging.com/software/texture-mtf-based-on-dead-leaves-target/>. [Accessed: 01-Jan-2017].
 - [29] P. D. Burns, "Refined Measurement of Digital Image Texture Loss," in *Proc. SPIE 8653, Image Quality and System Performance X*, 2013, p. 86530H.
 - [30] H. S. Malvar, L. He, and R. Cutler, "High-quality linear interpolation for demosaicing of Bayer-patterned color images," in *2004 IEEE International*

Conference on Acoustics, Speech, and Signal Processing, 2004, vol. 3, pp. 5–8.

- [31] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising with block-matching and 3D filtering,” in *Proc. SPIE 6064, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, 2006, pp. 606414-1-606414–12.
- [32] K. He, J. Sun, and X. Tang, “Guided Image Filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2010.
- [33] A. Haun and E. Peli, “Is image quality a function of contrast perception?,” in *Proc. SPIE 8651: Human Vision and Electronic Imaging XVIII*, 2013, pp. 86510C-1-86510C–5.
- [34] S. Daly, “The visible difference predictor: an algorithm for the assessment of image fidelity,” in *Digital Images and Human Vision*, A. B. Watson, Ed. The MIT Press, 1993, pp. 179–206.
- [35] S. Triantaphillidou, J. R. Jarvis, A. Psarrou, and G. Gupta, “Contrast detection in images of natural scenes,” *Signal Process. Image Commun.*, vol. 75, pp. 64–75, 2019.
- [36] D. A. Silverstein and J. E. Farrell, “The relationship between image fidelity and image quality,” in *Proc. 3rd IEEE International Conference on Image Processing*, 1996, vol. 1, pp. 881–884.
- [37] B. W. Keelan and H. Urabe, “ISO 20462: a psychophysical image quality measurement standard,” in *Proc. SPIE 5294: Image Quality and System Performance*, 2004, pp. 181–189.
- [38] R. E. Jacobson, “Approaches to total quality assessment of imaging systems,” *Inf. Serv. Use*, vol. 13, pp. 235–246, 1993.
- [39] E. W. Jin and B. W. Keelan, “Slider-adjusted softcopy ruler for calibrated image quality assessment,” *J. Electron. Imaging*, vol. 19, no. 1, pp. 011009-1-011009–12, 2010.
- [40] P. G. Engeldrum, “Psychometric Scaling : Avoiding the Pitfalls and Hazards,” in *PICS: Image Processing, Image Quality, Image Capture Systems Conference*, 2001, pp. 101–107.

-
- [41] P. D. Burns, "Image Quality Concepts," in *Handbook of Digital Imaging*, M. Kriss, Ed. Wiley, 2014, pp. 1–47.
- [42] R. E. Jacobson, "Image Quality Measurements: Necessity, Numbers and '.... nesses'." Plenary Talk: Digital Futures 2009: Image Physics & Psychophysics November 3rd 2009. The Institute of Physics, London, 2009.
- [43] O. H. Schade, "Image Quality: A Comparison of Photographic and Television Systems," *RCA Lab.*, 1975.
- [44] P. Vetter and A. Newen, "Varieties of cognitive penetration in visual perception," *Conscious. Cogn.*, vol. 27, pp. 62–75, 2014.
- [45] Z. Pylyshyn, "Is vision continuous with cognition? The case for cognitive impenetrability of visual perception.," *Behav. Brain Sci.*, vol. 22, no. 3, pp. 341–365, 1999.
- [46] B. W. Keelan, *Handbook of Image Quality, Characterisation and Prediction*. New York: Marcel Dekker, 2002.
- [47] I. E. Heynderickx and S. Bech, "Image quality assessment by expert and non-expert viewers," in *Proc. SPIE 4622, Human Vision and Electronic Imaging VII*, 2003, pp. 129–137.
- [48] M. Persson, "Subjective Image Quality Evaluation Using the Softcopy Quality Ruler Method," Lund University, 2014.
- [49] G. Deffner, M. Yuasa, M. Mckeon, and D. Arndt, "Evaluation of display-image quality: experts versus non-experts," in *1994 SID International Symposium, Digest of Technical Papers*, 1994, pp. 475–478.
- [50] DxOMark, "DxOMark: Find Scores and Reviews," 2017. [Online]. Available: <https://www.dxomark.com/>. [Accessed: 06-Feb-2017].
- [51] S. Yendrikhovskij, "Image Quality and Colour Categorisation," in *Colour Image Science*, L. W. Macdonald and M. R. Luo, Eds. John Wiley & Sons, 2002, pp. 392–418.
- [52] R. E. Jacobson, "An evaluation of image quality metrics," *J. Photogr. Sci.*, vol. 43,

pp. 7–16, 1995.

- [53] S. A. Klein, “Image quality and image compression: a psychophysicist’s viewpoint,” in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge: MIT Press, 1993, pp. 74–98.
- [54] S. Triantaphillidou, “Aspects of Image Quality in the Digitisation of Photographic Collections,” PhD Thesis, School of Communication and Creative Industries, University of Westminster, UK, 2001.
- [55] S. Yendrikhovskij, F. J. Blommaert, and H. de Ridder, “Towards perceptually optimal colour reproduction of natural scenes,” in *Colour Imaging Vision and Technology*, L. W. Macdonald and M. Ronnier Luo, Eds. Chichester: Wiley, 1999.
- [56] *ISO 20462-3:2012: Photography -- Psychophysical experimental methods for estimating image quality -- Part 3: Quality ruler method*. International Organization for Standardization (ISO), 2012.
- [57] K. H. Oh, S. Triantaphillidou, and R. E. Jacobson, “Perceptual image attribute scales derived from overall image quality assessments,” in *Proc SPIE 7242, Image Quality and System Performance VI*, 2009, p. 72420C.
- [58] K. H. Oh, S. Triantaphillidou, and R. E. Jacobson, “Scene classification with respect to image quality measurements,” in *Proc SPIE 7529, Image Quality and System Performance VII.*, 2010, p. 752908.
- [59] M. Orfanidou, S. Triantaphillidou, and E. Allen, “Predicting Image Quality using a Modular Image Difference Model,” in *Proc. SPIE 6808, Image Quality and System Performance V*, 2008, p. 68080F.
- [60] S. Triantaphillidou, E. Allen, and R. E. Jacobson, “Image Quality Comparison Between JPEG and JPEG2000. II. Scene Dependency, Scene Analysis, and Classification,” *J. Imaging Sci. Technol.*, vol. 51, no. 3, 2007.
- [61] C. N. Nelson, F. C. Eisen, and G. C. Higgins, “Effect Of Nonlinearities When Applying Modulation Transfer Techniques To Photographic Systems,” in *Proc. SPIE 13: Modulation Transfer Function*, 1968, pp. 127–134.
- [62] G. C. Higgins, “Methods for analyzing the photographic system, including the

-
- effects of nonlinearity and spatial frequency response,” *Photogr. Sci. Eng.*, vol. 15, no. 2, pp. 106–118, 1971.
- [63] S. E. Reichenbach, S. K. Park, and R. Narayanswamy, “Characterizing digital image acquisition devices,” *Opt. Eng.*, vol. 30, no. 2, pp. 170–177, 1991.
- [64] J. C. Feltz and M. A. Karim, “Modulation Transfer Function of Charge-Coupled Devices,” *Appl. Opt.*, vol. 29, pp. 717–722, 1990.
- [65] E. Allen, S. Triantaphillidou, and R. E. Jacobson, “Image Quality Comparison between JPEG and JPEG2000. I. Psychophysical Investigation,” *J. Imaging Sci. Technol.*, vol. 52, no. 1, pp. 1–15, 2008.
- [66] J. W. Glotzbach, R. W. Schafer, and K. Illgner, “A method of color filter array interpolation with alias cancellation properties,” in *Proc. IEEE 2001 International Conference on Image Processing*, 2001, vol. 1, pp. 141–144.
- [67] D. Menon, S. Andriani, and G. Calvagno, “Demosaicing with directional filtering and a posteriori decision,” *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 132–141, 2007.
- [68] M. Kriss, “Color Reproduction for Digital Cameras,” in *Handbook of Digital Imaging*, M. Kriss, Ed. Wiley Online Library, 2015, pp. 1–68.
- [69] L. Zhang, X. Wu, A. Buades, and X. Li, “Color demosaicking by local directional interpolation and nonlocal adaptive thresholding,” *J. Electron. Imaging*, vol. 20, no. 2, 2011.
- [70] Y. M. Lu, M. Karzand, and M. Vetterli, “Demosaicking by alternating projections: Theory and fast one-step implementation,” *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2085–2098, 2010.
- [71] Y. Hel-Or and D. Keren, *Demosaicing of color images using steerable wavelets*, vol. 206. HP Labs Israel, Technical Report HPL-2002-206R1, 2002.
- [72] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, “Color plane interpolation using alternating projections,” *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 997–1013, 2002.

-
- [73] R. C. Sumner, R. Burada, and N. Kram, “The Effects of Misregistration on the Dead Leaves Cross-Correlation Texture Blur Analysis,” in *Proc. IS&T Electronic Imaging: Image Quality and System Performance XIV*, 2017, pp. 121-129(9).
 - [74] J. B. Phillips, S. M. Coppola, E. W. Jin, Y. Chen, J. H. Clark, and T. A. Mauer, “Correlating objective and subjective evaluation of texture appearance with applications to camera phone imaging,” in *Proc. SPIE-IS&T Electronic Imaging 7242, Image Quality and System Performance VI*, 2009, pp. 724207-1-724207-11.
 - [75] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” in *International Conference on Computer Vision (ICCV)*, 1998, pp. 839-846.
 - [76] A. Buades, B. Coll, and J. Morel, “A Review of Image Denoising Algorithms, with a New One,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490-530, 2005.
 - [77] A. Foi, V. Katkovnik, and K. Egiazarian, “Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images,” *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1-17, 2007.
 - [78] C. Knaus and M. Zwicker, “Dual-domain image denoising,” in *Proc. IEEE International Conference on Image Processing*, 2013, no. 4, pp. 440-444.
 - [79] S. Roth and M. J. Black, “Fields of Experts: a framework for learning image priors,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, p. 10.1109/CVPR.2005.160.
 - [80] M. Elad and M. Aharon, “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 754-758, 2006.
 - [81] W. Dong, G. Shi, and X. Li, “Nonlocal image restoration with bilateral variance estimation: A low-rank approach,” *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700-711, 2013.
 - [82] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising with multi-layer perceptrons, part 1: comparison with existing algorithms and with bounds,” *J. Mach. Learn. Res.*, pp. 1-38, 2012.
 - [83] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole

-
- image restoration,” in *2011 IEEE International Conference on Computer Vision*, 2011, pp. 479–486.
- [84] X. Zhu and P. Milanfar, “Restoration for weakly blurred and strongly noisy images,” in *2011 IEEE Workshop on Applications of Computer Vision (WACV)*, 2011, pp. 103–109.
- [85] MathWorks, “Imsharpen: Sharpen image using unsharp masking,” 2017. [Online]. Available: <http://uk.mathworks.com/help/images/ref/imsharpen.html>. [Accessed: 05-Feb-2017].
- [86] G. Ramponi and A. Polesel, “Rational unsharp masking technique,” *J. Electron. Imaging*, vol. 7, no. 2, p. 333, 1998.
- [87] G. Deng, “A generalized unsharp masking algorithm,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1249–1261, 2011.
- [88] K. Kintali, “Adaptive Median Filter,” 2016. [Online]. Available: http://uk.mathworks.com/matlabcentral/fileexchange/30068-adaptive-median-filter--matlab-code-/content/AdaptiveMedianFilter_MATLAB_code/MALTAB_code/aMediantFilter_2D.m. [Accessed: 01-Feb-2017].
- [89] S. H. Kim and J. P. Allebach, “Optimal unsharp mask for image sharpening and noise removal,” *J. Electron. Imaging*, vol. 14, no. 2, 2005.
- [90] A. Polesel, G. Ramponi, and V. J. Mathews, “Image enhancement via adaptive unsharp masking,” *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 505–510, 2000.
- [91] E. S. L. Gastal and M. M. Oliveira, “Domain Transform for Edge-Aware Image and Video Processing,” in *ACM Transactions on Graphics (ToG)*, 2011, vol. 30, no. 4, p. 69.
- [92] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” in *Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2008*, 2008, vol. 27, no. 3, p. 67.
- [93] S. Ghosh and K. N. Chaudhury, “On Fast Bilateral Filtering Using Fourier Kernels,”

-
- IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 570–574, 2016.
- [94] N. Moroney, “Local color correction using non-linear masking,” *IS&T/SID 8th Color Imaging Conf.*, pp. 108–111, 2000.
 - [95] A. Capra, A. Castorina, S. Corchs, F. Gasparini, and R. Schettini, “Dynamic range optimization by local contrast correction and histogram image analysis,” *Dig. Tech. Pap. - IEEE Int. Conf. Consum. Electron.*, vol. 2006, pp. 309–310, 2006.
 - [96] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild, “Evaluating HDR Rendering Algorithms,” *ACM Trans. Appl. Percept.*, vol. 4, no. 2, p. 9, 2007.
 - [97] *ISO/IEC 10918-1:1994. Information technology -- Digital compression and coding of continuous-tone still images*. International Standardisation Organization (ISO) & International Electrotechnical Commission (IEC), 1994.
 - [98] G. K. Wallace, “The JPEG still picture compression standard,” *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
 - [99] *ISO/IEC 15444-1. Information technology -- JPEG 2000 image coding system: Core coding system*. International Organization for Standardization (ISO) & International Electrotechnical Commission (IEC), 2016.
 - [100] X. Min, G. Zhai, Z. Gao, and K. Gu, “Visual attention data for image quality assessment databases,” in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 894–897.
 - [101] H. Liu and I. Heynderickx, “Visual attention in objective image quality assessment: Based on eye-tracking data,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, 2011.
 - [102] E. Allen, “Image Quality Evaluation in Lossy Compressed Images,” PhD Thesis, Faculty of Media Arts and Design, University of Westminster, UK, 2017.
 - [103] J. Harel, C. Koch, and P. Perona, “Graph-Based Visual Saliency,” in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, 2006, pp. 545–552.
 - [104] E. M. Granger and K. N. Cupery, “An optical merit function (SQF) which correlates with subjective image judgments,” *Photogr. Sci. Eng.*, vol. 16, no. 3, pp. 221–230,

1972.

- [105] E. M. Biederman, “Photographic Korrespondent,” vol. 25 and 41, p. 103, 1967.
- [106] E. a. Fedorovskaya and H. De Ridder, “Subjective matters: from image quality to image psychology,” *SPIE Proc. Vol. 8651*, vol. 8651, pp. 865100-865100–11, 2013.
- [107] W. Fushikida, K. Schloss, K. Yokosawa, and S. Palmer, “Cross-Cultural Differences in Color Preference: Japan vs. the USA,” *J. Vis.*, vol. 9, no. 336, 2009.
- [108] L. A. Jones and H. R. Condit, “The Brightness scale of exterior scenes and the computation of correct photographic exposure,” *J. Opt. Soc. Am.*, vol. 31, no. 651, 1941.
- [109] M. H. R. Leyssen, S. Linsen, J. Sammartino, and S. E. Palmer, “Aesthetic Preference for Spatial Composition in Multiobject Pictures,” *Iperception.*, vol. 3, no. 1, pp. 25–49, 2012.
- [110] *ISO 20462-1:2005: Photography — Psychophysical experimental methods for estimating image quality — Part 1: Overview of Psychophysical Elements*. International Organization for Standardization (ISO), 2005.
- [111] C. J. Bartleson, “The combined influence of sharpness and graininess on the quality of color prints,” *J. Photogr. Sci.*, vol. 30, pp. 33–38, 1982.
- [112] J. F. Sawyer, “Effect of graininess and sharpness on perceived print quality,” in *Photographic Image Quality Symposium*, 1980, pp. 222–231.
- [113] A. M. Haun and E. Peli, “Complexities of complex contrast,” in *Proc SPIE 8292, Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications*, 2012.
- [114] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [115] W. S. Geisler, “Visual Perception and the Statistical Properties of Natural Scenes,” *Annu. Rev. Psychol.*, vol. 59, pp. 167–192, 2008.
- [116] E. P. Simoncelli and B. A. Olshausen, “Natural image statistics and neural

-
- representation,” *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–216, 2001.
- [117] B. Cooper, H. Sun, and B. Lee, “Psychophysical and physiological responses to gratings with luminance and chromatic components of different spatial frequencies,” *J. Opt. Soc. Am. A*, vol. 28, no. 2, pp. A314–A323, 2012.
- [118] E. Peli, “Contrast in complex images,” *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 2032–2040, 1990.
- [119] S. Triantaphillidou, R. E. Jacobson, and A. M. Ford, “Preferred Tone Reproduction of Images on Soft Displays,” in *International Conference on the Physics of Semiconductors (ICPS)*, 1998, pp. 205–208.
- [120] R. W. G. Hunt, I. T. Pitt, and P. C. Ward, “The Tone Reproduction of Colour Photographic Materials,” *J. Photogr. Sci.*, vol. 17, pp. 198–204, 1969.
- [121] A. C. Bovik, *Handbook of Image and Video Processing*, 2nd ed. Academic Press Ltd., 2010.
- [122] B. Bringier, N. Richard, M. C. Larabi, and C. Fernandez-Maloigne, “No-reference perceptual quality assessment of colour image,” in *Proceedings of the 14th IEEE European Signal Processing Conference (EUSIPCO 2006)*, 2006.
- [123] A. M. Haun and E. Peli, “Perceived contrast in complex images,” *J. Vis.*, vol. 13, no. 3, pp. 1–21, 2013.
- [124] A. M. Haun and E. Peli, “Measuring the Perceived Contrast of Natural Images,” *SID Symp. Dig. Tech. Pap.*, vol. 42, pp. 302–304, 2011.
- [125] S. Triantaphillidou, J. Jarvis, G. Gupta, and H. Rana, “Defining human contrast sensitivity and discrimination from complex imagery,” in *Proc. SPIE 8901, Optics and Photonics for Counterterrorism, Crime Fighting and Defence IX*, 2013, pp. 89010C-1-89010C-12.
- [126] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, 2006.
- [127] J. A. J. Roufs, V. J. Koselka, and A. A. Tongeren, “Global brightness contrast and the effect on perceptual image quality,” in *Proc SPIE 2179, Human Vision, Visual*

Processing, and Digital Display V; 1994, pp. 80–89.

- [128] C. A. Poynton, “Gamma and its disguises: the non-linear mappings of intensity in perception, CRTs, film and video,” *J. Soc. Mot. Pict. Telivis. Engrs.*, pp. 1099–1108, 1993.
- [129] *ISO 14524:2009 Photography - Electronic still-picture cameras - Methods for measuring opto-electronic conversion functions (OECFs)*. International Organization for Standardization (ISO), 2009.
- [130] Image Engineering, “TE264 OECF 20 Chart (following ISO 14524/15739) contrast 10.000:1, 100.000:1, 1.000.000:1,” 2019. [Online]. Available: <https://www.image-engineering.de/products/charts/all/581-te264>. [Accessed: 03-Mar-2019].
- [131] R. S. Berns, “Methods for characterizing CRT displays,” *Displays*, vol. 16, no. 4, pp. 173–182, 1996.
- [132] *BS EN 61966-2-1:2000 Multimedia systems and equipment — Colour measurement and management — Part 2-1: Colour management - Default RGB colour space - sRGB*. International Organization for Standardization (ISO), 2000.
- [133] D. J. Field and N. Brady, “Visual Sensitivity, Blur and the Sources of Variability in the Amplitude Spectra of Natural Scenes,” *Vis. Res.*, vol. 37, no. 23, pp. 3367–3383, 1997.
- [134] K. Biedermann and Y. Feng, “Quality criteria for photographic images and lenses,” in *Photographic and Electronic Image Quality*, 1984, pp. 32–37.
- [135] L. Macdonald, “Framework for an Image Sharpness Management System,” in *Proc. IS&T/SID, The Seventh Color Imaging Conference: Color Science, Systems, and Applications*, 1999, pp. 75–79.
- [136] G. C. Higgins, “Image Quality Criteria,” *J. Appl. Photogr. Eng.*, vol. 3, no. 2, pp. 53–60, 1977.
- [137] G. P. Corey, M. J. Clayton, and K. N. Cupery, “Scene Dependence of Image Quality,” *Photogr. Sci. Eng.*, vol. 27, pp. 9–13, 1983.
- [138] S. Bouzit and L. W. Macdonald, “Assessing the enhancement of image sharpness,”

-
- in *Proc. SPIE 6059, Image Quality and System Performance III*, 2006, pp. 605904-1-605904-10.
- [139] S. Bouzit and L. W. MacDonald, "Sharpness Enhancement through Spatial Frequency Decomposition," in *Proc IS&T, PICS 2001: Image Processing, Image Quality, Image Capture Systems Conference*, 2001, pp. 377-381.
- [140] G. M. Johnson and M. D. Fairchild, "Sharpness Rules," in *Proc. IS&T, Eighth Color Imaging Conference: Color Science and Engineering Systems Technologies, Applications 8*, 2000, pp. 1-28.
- [141] N. Koren, *Imatest Documentation* © 2009. Imatest LLC, 2009.
- [142] C. Loebich, D. Wueller, B. Klingen, and A. Jaeger, "Digital Camera Resolution Measurement Using Sinusoidal Siemens Stars," in *Proc. SPIE 6502, Digital Photography III*, 2007.
- [143] S. Bouzit and L. W. Macdonald, "Does Sharpness Affect the Reproduction of Colour Images?," in *Proc. SPIE 4421, 9th Congress of the International Colour Association*, 2002, pp. 902-905.
- [144] J. Buzzi and F. Guichard, "Uniqueness of Blur Measure," in *2004 IEEE International Conference on Image Processing (ICIP)*, 2004, pp. 2985-2988.
- [145] LaserSoft Imaging, "SilverFast Resolution Target (USAF 1951) by LaserSoft Imaging," 2019. [Online]. Available: <https://www.silverfast.com/show/resolution-target/en.html>. [Accessed: 06-Aug-2019].
- [146] U. Artmann, "Image quality assessment using the dead leaves target : experience with the latest approach and further investigations," in *Proc. SPIE 9404, Digital Photography XI*, 2015, pp. 1-15.
- [147] E. Heynacher and F. Kober, *Zeiss Information No 51: Resolving Power and Contrast*. Oberkochen: Carl Zeiss, 1976.
- [148] R. A. Jones and E. C. Yeadon, "Determination of the spread function from noisy edge scans," *Photogr. Sci. Eng.*, vol. 13, no. 4, pp. 200-204, 1969.
- [149] R. A. Jones, "An automated technique for deriving MTF's from edge traces,"

Photogr. Sci. Eng., vol. 13, no. 2, pp. 102–106, 1967.

- [150] J. C. Dainty, “Methods of Measuring the Modulation Transfer Function of Photographic Emulsions,” *Opt. Acta Int. J. Opt.*, vol. 18, no. 11, pp. 795–813, 1971.
- [151] Image Engineering, “TE265 - Dead Leaves,” 2019. [Online]. Available: <https://www.image-engineering.de/products/charts/all/582-te265>. [Accessed: 01-Jan-2019].
- [152] Imatest, “Documentation: Log F-Contrast,” 2019. [Online]. Available: http://www.imatest.com/docs/log_f_cont/. [Accessed: 17-Mar-2019].
- [153] *ANSI-PH2.29-1977. Method of Measuring the Photographic Modulation Transfer Function of Continuous-Tone, Black-and-White Photographic Films*. New York, USA: American National Standards Institute (ANSI), 1977.
- [154] R. L. Lamberts, “Sine-wave response techniques in photographic printing,” *J. Opt. Soc. Am. A*, vol. 51, p. 982, 1961.
- [155] J. W. Coltman, “The specification of imaging properties by response to a sine wave input,” *J. Opt. Soc. Am. A*, vol. 44, p. 468, 1954.
- [156] R. B. Fagard-Jenkin, R. E. Jacobson, and N. R. Axford, “A Novel Approach to the Derivation of Expressions for Geometrical MTF in Sampled Systems,” *Soc. Imaging Sci. Technol. Image Process. Image Qual. Image Capture, Syst. Conf.*, pp. 225–230, 1999.
- [157] E. C. Yeadon, R. A. Jones, and J. T. Kelly, “Confidence Limits for Individual Modulation Transfer Function Measurements based on the Phase Transfer Function,” *Photogr. Sci. Eng.*, vol. 14, no. 2, pp. 153–156, 1970.
- [158] P. D. Burns, “Estimation Error in Image Quality Measurements,” in *Proc. SPIE 7867, Image Quality and System Performance VIII*, 2011, p. 78670H.
- [159] U. Artmann and D. Wueller, “Differences of digital camera resolution metrology to describe noise reduction artifacts,” in *Proc. SPIE 7529: Image Quality and System Performance VII*, 2010, pp. 1–12.
- [160] U. Artmann, “Linearization and Normalization in Spatial Frequency Response

-
- Measurement,” in *Proc. IS&T Electronic Imaging, Image Quality and System Performance XIII*, 2016, pp. 1–6.
- [161] P. D. Burns and J. Martinez Bauza, “Intrinsic camera resolution measurement,” in *Proc. SPIE 9396, Image Quality and System Performance XII*, 2015, p. 939609.
- [162] H. Kurihara, T.; Aoki, N.; Kobayashi, “Analysis of sharpness increase by image noise,” *J. Imaging Sci. Technol.*, vol. 55, no. 3, pp. 30504–1, 2011.
- [163] X. Wan, N. Aoki, and H. Kobayashi, “Improving the Perception of Image Sharpness Using Noise Addition,” *Bull. Soc. Photogr. Imag. Japan.*, vol. 24, no. 2, pp. 19–26, 2014.
- [164] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt Jr., and J. M. Boone, *The Essential Physics of Medical Imaging*, Second. Philadelphia, USA: Lippincott Williams & Wilkins, 2002.
- [165] R. B. Jenkin, *Private communication*. 2017.
- [166] R. B. Jenkin, “Noise, sharpness, resolution and information,” in *The Manual of Photography: Photographic and Digital Imaging*, 10th ed., E. Allen and S. Triantaphillidou, Eds. Focal Press, 2011.
- [167] J. E. Farrell and B. A. Wandell, “Image Systems Simulation,” in *Handbook of Digital Imaging*, M. Kriss, Ed. Wiley, 2015, pp. 1–28.
- [168] *ISO 15739:2017, Photography — Electronic still-picture imaging — Noise measurements*. International Organization for Standardization (ISO), 2017.
- [169] K. Topfer and R. E. Jacobson, “The Relationship Between Objective and Subjective Image Quality Criteria,” *J. Inf. Rec. Mater.*, vol. 21, pp. 5–27, 1993.
- [170] P. D. Welch, “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoust.*, vol. 15, pp. 70–73, 1967.
- [171] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.
- [172] Imatest, “Imatest Documentation: Shannon Information Capacity,” 2019. [Online].

Available: <http://www.imatest.com/docs/shannon/>. [Accessed: 26-Mar-2019].

- [173] The Society for Imaging Science and Technology (IS&T), *Handbook of Photographic Science and Engineering.*, 2nd ed. The Society for Imaging Science and Technology (IS&T), 1997.
- [174] I. E. Abdou and N. J. Dusaussay, "Survey of Image Quality Measurements," in *Proceedings of 1986 IEEE ACM Fall Joint Computer Conference*, 1986, pp. 71–78.
- [175] B. Rodricks and K. Venkataraman, "First Principles' Imaging Performance Evaluation of CCD- and CMOS-based Digital Camera Systems," in *Proc. SPIE 5678, Digital Photography, 59 - Sensor and Camera Characterization*, 2005, pp. 59–74.
- [176] B. W. Keelan, "Imaging Applications of Noise Equivalent Quanta," in *Proc. IS&T Electronic Imaging 2016, Image Quality and System Performance XIII*, 2016, pp. 1–7.
- [177] P. D. Burns, "Signal-to-noise ratio analysis of charge-coupled device imagers," in *Proc. SPIE 1242, Charge-Coupled Devices and Solid State Optical Sensors*, 1990, pp. 187–194.
- [178] B. W. Keelan, "Objective and subjective measurement and modeling of image quality: a case study," in *Proc. SPIE 7798: Applications of Digital Image Processing XXXIII*, 2010, pp. 779815-1-779815–9.
- [179] H. H. Barrett, "NEQ: its progenitors and progeny," in *Proc. SPIE 7263, Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment*, 2009, pp. 72630F-1-72630F–7.
- [180] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A*, vol. 12, no. 5, pp. 834–852, 1995.
- [181] *ICRU Report 54: Medical Imaging - The Assessment of Image Quality*. Bethesda, MD, USA: International Commission on Radiation Units and Measurements (ICRU), 1996.
- [182] P. G. J. Barten, "Physical model for the contrast sensitivity of the human eye," in

Proc. SPIE/IS&T Electronic Imaging: Science and Technology 1666, Human Vision, Visual Processing, and Digital Display III, 1992, pp. 57–72.

- [183] P. G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. Bellingham, WA: SPIE press, 1999.
- [184] S. Daly, “A visual model for optimizing the design of image processing algorithms,” in *Proc. IEEE 1st International Conference on Image Processing*, 1994, vol. 2, pp. 16–20.
- [185] J. Lubin, “The Use of Psychophysical Data and Models in the Analysis of Display System Performance,” in *Digital Images and Human Vision*, A. Watson, Ed. MIT Press, 1993, pp. 163–178.
- [186] P. G. J. Barten, “The Square Root Integral (SQRI): a New Metric to Describe the Effect of Various Display Parameters on Image Quality,” in *Proc. SPIE 1077: Human Vision, Visual Processing, and Digital Display*, 1989, pp. 73–82.
- [187] S. Triantaphillidou, J. Jarvis, and G. Gupta, “Contrast sensitivity and discrimination of complex scenes,” in *Proc. SPIE 8653, Image Quality and System Performance X*, 2013, p. 86530C.
- [188] G. M. Johnson and M. D. Fairchild, “A top down description of S-CIELAB and CIEDE2000,” *Color Res. Appl.*, vol. 28, no. 6, pp. 425–435, 2003.
- [189] J. A. Movshon and L. Kiorpes, “Analysis of the development of spatial sensitivity in monkey and human infants,” *J. Opt. Soc. Am.*, vol. 5, no. 12, pp. 2166–2172, 1988.
- [190] M. A. Georgeson and G. D. Sullivan, “Contrast Constancy: Deblurring in Human Vision by Spatial Frequency Channels,” *J. Physiol.*, vol. 1, no. 4, pp. 627–656, 1975.
- [191] O. H. Schade, “Optical and photoelectric analog of the eye,” *J. Opt. Soc. Am.*, vol. 46, pp. 721–739, 1956.
- [192] J. A. J. Roufs, “Dynamic properties of vision – V1. Stochastic threshold fluctuations and their effect on flash-to-flicker sensitivity ratio,” *Vision Res.*, vol. 14, pp. 871–888, 1974.

-
- [193] G. E. Legge, D. Kersten, and A. E. Burgess, "Contrast discrimination in noise.," *J. Opt. Soc. Am. A*, vol. 4, no. 2, pp. 391–404, 1987.
 - [194] V. Laparra, J. Muñoz-Marí, and J. Malo, "Divisive normalization image quality metric revisited," *J. Opt. Soc. Am. A JOSA-A, Opt. image Sci. Vis.*, vol. 27, no. 4, pp. 852–864, 2010.
 - [195] R. B. Jenkin, S. Triantaphillidou, and M. A. Richardson, "Effective Pictorial Information Capacity as an Image Quality Metric," in *Proc. SPIE 6494, Image Quality and System Performance IV*, 2007, p. 64940O.
 - [196] A. van Meeteren, "Visual aspects of image intensification," PhD Thesis, University of Utrecht, The Netherlands, 1973.
 - [197] H. L. Snyder, "Image Quality and Observer Performance," in *Perception of Displayed Information*, L. M. Biberman, Ed. New York: Plenum Press, 1973, pp. 87–118.
 - [198] R. G. Gendron, "An improved objective method for rating picture sharpness: CMT acutance," *J. Soc. Motion Pict. Telev. Eng.*, vol. 82, pp. 1009–1012, 1973.
 - [199] E. M. Crane, "Acutance and Granulance," in *Proc. SPIE 310, Image Quality*, 1981, pp. 125–132.
 - [200] B. E. Rogowitz, T. N. Pappas, and J. P. Allebach, "Human Vision and Electronic Imaging," *J. Electron. Imaging*, vol. 10, no. 1, pp. 10–19, 2001.
 - [201] S. Triantaphillidou, E. W. S. Fry, V. Sanchis-Jurado, and A. Pons, "Image Quality Loss and Compensation for Visually Impaired Observers," in *Proc IS&T Electronic Imaging, Image Quality and System Performance XV*, 2018, pp. 365-1-365–6(6).
 - [202] Z. Wang, "Applications of objective image quality assessment methods," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, 2011.
 - [203] R. Dosselmann and X. D. Yang, "A Formal Assessment of the Structural Similarity Index," Technical Report TR-CS 2008-2, Department of Computer Science, University of Regina, Canada, 2008.
 - [204] M. Pedersen and J. Y. Hardeberg, "Survey of full-reference image quality metrics,"

-
- GCIS'2009 Glob. Congr. Intell. Syst.*, no. 5, pp. 1–74, 2009.
- [205] K. Seshadrinathan *et al.*, “Image Quality Assessment,” in *Essential Guide of Image Processing*, Elsevier, 2009, pp. 553–589.
- [206] X. Gao, W. Lu, D. Tao, and X. Li, “Image quality assessment and human visual system,” in *Proc. SPIE 7744, Visual Communications and Image Processing 2010*, 2010, pp. 77440Z–77440Z–10.
- [207] K. Seshadrinathan and A. C. Bovik, “Automatic prediction of perceptual quality of multimedia signals-a survey,” *Multimed. Tools Appl.*, vol. 51, pp. 163–186, 2011.
- [208] M. Pedersen and J. Y. Hardeberg, “Full-Reference Image Quality Metrics: Classification and Evaluation,” *Found. Trends Comput. Graph. Vis.*, vol. 7, no. 1, pp. 1–80, 2011.
- [209] M. Pedersen, “Image quality metrics for the evaluation of printing workflows,” PhD Thesis, Faculty of Mathematics and Natural Sciences, University of Oslo, Norway, 2011.
- [210] A. Lahoulou, M. C. Larabi, A. Beghdadi, E. Viennet, and A. Bouridane, “Knowledge based Taxonomic Scheme for Full Reference Objective Image Quality Measurement Models,” *J. Imaging Sci. Technol.*, vol. 60, no. 6, pp. 64–78, 2016.
- [211] T. Eerola *et al.*, “Full Reference Printed Image Quality: Measurement Framework and Statistical Evaluation,” *J. Imaging Sci. Technol.*, vol. 54, no. 1, pp. 1–13, 2010.
- [212] A. K. Moorthy and A. C. Bovik, “Visual quality assessment algorithms: what does the future hold?,” *Multimed. Tools Appl.*, vol. 51, no. 2, pp. 675–696, 2010.
- [213] J. J. Gallimore, “Review of Psychophysically Based Image Quality Metrics,” Wright State University, Defense Technical Information Center, 1991.
- [214] Z. Wang and A. Bovik, “Reduced and No-Reference Image Quality Assessment,” *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, 2011.
- [215] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albrechtsen, “Image quality metrics for the evaluation of print quality,” in *Proc. SPIE 7867, Image Quality and System Performance VIII*, 2011, pp. 786702–786702–19.

-
- [216] D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," *ISRN Signal Process.*, vol. 2013, pp. 1–53, 2013.
 - [217] R. Soundararajan and A. C. Bovik, "Survey of information theory in visual quality assessment," *Signal, Image Video Process.*, vol. 7, no. 3, pp. 391–401, 2013.
 - [218] M. P. Eckert and A. P. Bradley, "Perceptual quality metrics applied to still image compression," *Signal Processing*, vol. 70, pp. 177–200, 1998.
 - [219] Z. Wang, "Objective Image / Video Quality Measurement - A Literature Survey," in *EE 381k: multidimensional digital signal processing*, 1998.
 - [220] W. Osberger, "Perceptual Vision Models for Picture Quality Assessment and Compression Applications," PhD Thesis, Queensland University of Technology, Australia, 1999.
 - [221] A. M. Eskicioglu, "Quality measurement for monochrome compressed images in the past 25 years," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2000, pp. 1907–1910.
 - [222] A. Chalmers, S. Daly, K. Myszkowskin, and T. Troscianko, "Image Quality Metrics," *SIGGRAPH Course #44*, 2000.
 - [223] S. Winkler, "Vision Models and Quality Metrics for Image Processing Applications," PhD Thesis, Departmente d'Electricite, Ecole Polytechnique Federale de Lausanne, Switzerland, 2000.
 - [224] K. Seshadrinathan and A. C. Bovik, "New vistas in image and video quality assessment," in *Proc SPIE 6492, Human Vision and Electronic Imaging XII*, 2007, pp. 649202–649202–13.
 - [225] University of Texas Laboratory for Image & Video Engineering, "LIVE Image Quality Assessment Database," 2017. [Online]. Available: <http://live.ece.utexas.edu/research/quality/subjective.htm>. [Accessed: 31-Jul-2017].
 - [226] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," 2008. [Online]. Available: <http://live.ece.utexas.edu/research/quality>. [Accessed: 13-Jun-2017].

-
- [227] P. Le Callet and F. Autrusseau, "Subjective quality assessment - IVC database," 2005. [Online]. Available: <http://www2.irccyn.ec-nantes.fr/ivcdb/>. [Accessed: 13-Jun-2016].
 - [228] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging*, vol. 19, no. 1, p. 011006, 2010.
 - [229] X. Liu, "CID : IQ - A new image quality database," PhD Thesis, Department of Computer Science and Media Technology, Gjøvik University College, Norway, 2013.
 - [230] X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID:IQ - A New Image Quality Database," in *Proc. 6th International Conference on Image and Signal Processing (ICISP)*, 2014, pp. 193–202.
 - [231] U. Engelke, M. Kusuma, H. J. Zepernick, and M. Caldera, "Reduced-reference metric design for objective perceptual quality assessment in wireless imaging," *Signal Process. Image Commun.*, vol. 24, no. 7, pp. 525–547, 2009.
 - [232] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, 2007.
 - [233] D. M. Chandler and S. S. Hemami, "Cornell-CVL A57 Database," 2007. [Online]. Available: <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
 - [234] University of Texas Laboratory for Image & Video Engineering, "LIVE In the Wild Image Quality Challenge Database," 2017. [Online]. Available: <http://live.ece.utexas.edu/research/ChallengeDB/index.html>. [Accessed: 31-Jul-2017].
 - [235] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *Image Process. IEEE Trans.*, vol. 15, no. 11, pp. 3441–3452, 2006.
 - [236] D. Jayaraman, A. Anish Mittal, K. Moorthy, and A. C. Bovik, "Objective Image Quality Assessment of Multiply Distorted Images," in *Proceedings of the Forty*

-
- Sixth IEEE Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 1058–1697.
- [237] N. Ponomarenko *et al.*, “TID2008—A database for evaluation of full-reference visual quality assessment metrics,” *Adv. Mod. Radioelectron.*, vol. 10, pp. 30–45, 2009.
- [238] N. Ponomarenko, “Tampere Image Database 2013 (TID2013), Version 1.0,” 2013. [Online]. Available: <http://www.ponomarenko.info/tid2013.htm>. [Accessed: 13-Jun-2016].
- [239] N. Ponomarenko *et al.*, “A New Color Image Database TID2013: Innovations and Results,” in *International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, 2013, pp. 402–413.
- [240] A. Ciancio, A. L. N. Targino da Costa, E. A. B. da Silva, A. Said, R. Samadani, and P. Obrador, “No-reference blur assessment of digital pictures based on multifeature classifiers,” *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 64–75, 2011.
- [241] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. Parvez Sazzad, “MICT-Toyama Image Quality Evaluation Database,” 2011. [Online]. Available: <http://mict.eng.u-toyama.ac.jp/mictdb.html>. [Accessed: 01-Dec-2015].
- [242] H. R. Sheikh, A. C. Bovik, L. Cormack, and Z. Wang, “LIVE Image Quality Assessment Database Release 1: Subjective database for JPEG Readme File,” 2003. [Online]. Available: <https://live.ece.utexas.edu/research/quality/JPEG/readme.txt>. [Accessed: 17-May-2019].
- [243] P. G. Engeldrum, “A short image quality model taxonomy,” *J. Imaging Sci. Technol.*, vol. 48, no. 2, pp. 160–165, 2004.
- [244] E. Crane, “An Objective Method of Rating Picture Sharpness: SMT Acutance,” *SMPTE Motion Imaging J.*, vol. 73, no. 8, pp. 643–647, 1964.
- [245] C. Carlson and R. Cohen, “A simple psychophysical model for predicting the visibility of displayed information,” in *Proc. Society of Information Display (SID)*, 1980, vol. 21, pp. 229–245.
- [246] H. L. Snyder, “Modulation Transfer Function Area as a Measure of Image Quality,” in *Visual Search Symposium Committee on Vision*, 1973, pp. 93–105.

-
- [247] P. G. J. Barten, "The SQRI Method: A New Method for the Evaluation of Visible Resolution on a Display," in *Proc. SID*, 1987, vol. 28, no. 3, pp. 253–262.
 - [248] C. N. Nelson, "Image Sharpness Criteria," *J. Opt. Soc. Amer.*, vol. 63, p. 1289, 1973.
 - [249] J. H. D. M. Westerlink and J. A. J. Roufs, "Subjective image quality as a function of viewing distance, resolution and picture size," *SMPTE Motion Imaging J.*, vol. 98, no. 2, pp. 113–119, 1989.
 - [250] P. G. J. Barten, "Evaluation of subjective image quality with the square-root integral method," *J. Opt. Soc. Am. A*, vol. 7, no. 10, p. 2024, 1990.
 - [251] R. B. Jenkin, R. E. Jacobson, and M. A. Richardson, "Use of the First Order Wiener Kernel Transform in the Evaluation of SQRI and PIC Quality Metrics for JPEG Compression," in *Proc. SPIE 5294, Image Quality and System Performance*, 2004, pp. 60–70.
 - [252] R. B. Jenkin and M. A. Richardson, "Comparison between the effective pictorial information capacities of JPEG 6b and 2000," in *Proc. SPIE 5823, Opto-Ireland 2005*, 2005, pp. 13–19.
 - [253] J. H. Altman and H. J. Zweig, "Effects of Spread Function on the Storage of Information on Photographic Emulsions," *Photogr. Sci. Eng.*, vol. 7, no. 3, pp. 173–177, 1963.
 - [254] K. H. Oh, S. Triantaphillidou, and R. E. Jacobson, "Device-dependent scene-dependent quality predictions using effective pictorial information capacity," in *Proc. SPIE 7867, Image Quality and System Performance VIII*, 2011.
 - [255] D. J. Heeger and P. C. Teo, "A model of perceptual image fidelity," in *IEEE International Conference on Image Processing*, 1995, vol. 2, pp. 343–345.
 - [256] T. Mitsa and K. L. Varkur, "Evaluation of Contrast Sensitivity Functions for the Formulation of Quality Measures Incorporated in Halftoning Algorithms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1993, pp. 301–304.
 - [257] R. N. Shepherd, "Metric structures in ordinal data," *J. Math. Psychol.*, vol. 3, pp.

287–315, 1966.

- [258] P. E. Green, F. J. Carmone, and S. M. Smith, *Multidimensional scaling: concepts and applications*. Boston: Allyn and Bacon, 1989.
- [259] L. A. Olzak and J. P. Thomas, “Seeing spatial patterns,” in *Handbook of perception and human performance, Vol. 1: Sensory processes*, L. Kaufman and J. P. Thomas, Eds. New York: John Wiley and Sons, 1986.
- [260] H. de Ridder, “Subjective evaluation of scale-space image coding,” in *Proc. SPIE 1453, Human Vision, Visual Processing, and Digital Display II*, 1991, pp. 31–42.
- [261] H. De Ridder, “Minkowski-metrics as a combination rule for digital-image-coding impairments,” in *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, 1992, pp. 16–26.
- [262] *ISO 11664-4:2011 Colorimetry: Part 4: CIE 1976 L*a*b* Colour space*. Commission Internationale de l’Eclairage (CIE), 2011.
- [263] B. Keelan, “Predicting Multivariate Image Quality from Individual Perceptual Attributes,” in *Proc. IS&T 2002 PICS Conference*, 2002, pp. 82–87.
- [264] S. Daly, “Visible differences predictor: an algorithm for the assessment of image fidelity,” in *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, 1992, pp. 179–206.
- [265] X. Zhang and B. A. Wandell, “A spatial extension of CIELAB for digital color image reproduction,” *J. Soc. Inf. Disp.*, vol. 5, no. 1, p. 61, 1997.
- [266] X. Zhang, J. E. Farrell, and B. A. Wandell, “Applications of a Spatial Extension to CIELAB,” in *Proc. IS&T & SPIE Electronic Imaging 3025*, 1997, pp. 154–157.
- [267] X. Zhang, D. A. Silverstein, J. E. Farrell, and B. A. Wandell, “Color image quality metric S-CIELAB and its application on halftone texture visibility,” in *Proc. IEEE COMPCON 97. Digest of Papers*, 1997.
- [268] P. Teo and D. J. Heeger, “Perceptual Image Distortion,” in *Proc. IEEE 1st International Conference on Image Processing (ICIP)*, 1994, pp. 982–986.
- [269] M. Miyahara, K. Kotani, and V. R. Algazi, “Objective picture quality scale (PQS)

-
- for image coding,” *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1215–1226, 1998.
- [270] C. H. Chou and Y.-C. Li, “A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, 1994.
- [271] S. A. Karunasekera and N. G. Kingsbury, “A distortion measure for image artifacts based on human visual sensitivity,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-94*, 1994, pp. 117–120.
- [272] J. Lubin, “A visual discrimination model for imaging system design and evaluation,” in *Vision Models for Target Detection and Recognition: In Memory of Arthur Menendez.*, 1995, pp. 245–283.
- [273] J. L. Mannos and D. J. Sakrison, “The Effects of a Visual Fidelity Criterion on the Encoding of Images,” *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 525–536, 1974.
- [274] R. J. Safranek and J. D. Johnston, “A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, 1989, pp. 1945–1948.
- [275] G. M. Johnson and M. D. Fairchild, “On contrast sensitivity in an image difference model,” in *Proc. IS&T PICS Conference*, 2002, pp. 18–23.
- [276] M. D. Fairchild and G. M. Johnson, “iCAM framework for image appearance, differences, and quality,” *J. Electron. Imaging*, vol. 13, no. 1, p. 126, 2004.
- [277] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, 2000.
- [278] M. R. Luo, G. Cui, and B. Rigg, “The development of the CIE 2000 colour-difference formula: CIEDE2000,” *Color Res. Appl.*, vol. 7, no. 13, pp. 340–350, 2001.
- [279] M. D. Fairchild and G. M. Johnson, “Meet iCAM: A next-generation color appearance model,” in *Proc. IS&T/SID 10th Color Imaging Conference*, 2002, pp. 33–38.

-
- [280] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, 2012.
 - [281] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.
 - [282] Z. Wang and A. C. Bovik, "Why is image quality assessment so difficult?," in *IEEE International Conference on Acoustics, Speech, & Signal Processing*, 2002, pp. 3313–3316.
 - [283] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-Scale Structural Similarity For Image Quality Assessment," in *Proc. IEEE Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, vol. 2, pp. 1398–1402.
 - [284] D. M. Rouse and S. S. Hemami, "Analyzing the Role of Visual Structure in the Recognition of Natural Image Content with Multi-Scale SSIM," in *Proc. SPIE 6806, Human Vision and Electronic Imaging XIII.*, 2008, p. 680615.
 - [285] Z. Wang and E. P. Simoncelli, "Translation Insensitive Image Similarity in Complex Wavelet Domain," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, vol. 2, pp. 573–576.
 - [286] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 422–429, 2006.
 - [287] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Networks*, vol. 21, no. 3, pp. 515–9, 2010.
 - [288] X. Gao, T. Wang, and J. Li, "A Content-Based Image Quality Metric," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, Springer, 2005, pp. 231–240.
 - [289] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An improved full-reference image quality metric based on structure compensation," in *Proc. IEEE 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012, pp. 1–6.
 - [290] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality

-
- measures,” *J. Electron. Imaging*, vol. 11, no. 2, pp. 206–223, 2002.
- [291] R. Gupta, D. Bansal, and C. Singh, “A Survey on Various Objective Image Quality Assessment Techniques,” *Int. J. Eng. Tech. Res.*, vol. 2, no. 7, pp. 99–104, 2014.
- [292] B. Girod, “What’s Wrong with Mean-squared Error?,” in *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, 1993, pp. 207–220.
- [293] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [294] Z. Cui, Z. Gan, G. Tang, F. Liu, and X. Zhu, “Simple and Effective Image Quality Assessment Based on Edge Enhanced Mean Square Error,” in *Proc. IEEE Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, 2014, pp. 1–5.
- [295] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, “Two New Full-Reference Quality Metrics based on HVS,” in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, 2006, pp. 2–5.
- [296] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, “On Between-coefficient Contrast Masking of DCT Basis Functions,” in *Proc. Third International Workshop on Video Processing and Quality Metrics*, 2007, vol. 4, pp. 1–4.
- [297] K. Gu, G. Zhai, X. Yang, L. Chen, and W. Zhang, “Nonlinear additive model based saliency map weighting strategy for image quality assessment,” in *Proc. IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 313–318.
- [298] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, “Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric,” in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2007, vol. 2, pp. 169–172.
- [299] Q. Li and Z. Wang, “Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation,” *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 202–211, 2009.

-
- [300] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation.," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–89, 2012.
- [301] K. H. Thung and P. Raveendran, "A survey of image quality measures," in *Proc. IEEE International Conference for Technical Postgraduates (TECHPOS)*, 2009, pp. 1–4.
- [302] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [303] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–64, 2011.
- [304] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [305] H. Tang and N. Joshi, "Learning a Blind Measure of Perceptual Image Quality," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 305–312.
- [306] T. H. Falk, Y. Guo, and W. Y. Chan, "Improving robustness of image quality measurement with degradation classification and machine learning," in *IEEE 2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, 2007, pp. 503–507.
- [307] D. Ghadiyaram and A. C. Bovik, "Blind image quality assessment on real distorted images using deep belief nets," in *IEEE Global Conference on Signal and Information Processing*, 2014, pp. 946–950.
- [308] Y. Lv, G. Jiang, M. Yu, H. Xu, F. Shao, and S. Liu, "Difference of Gaussian statistical features based blind image quality assessment: A deep learning approach," in *IEEE International Conference on Image Processing*, 2015, pp. 2344–2348.

-
- [309] H. Tang, N. Joshi, and A. Kapoor, “Blind Image Quality Assessment Using Semi-supervised Rectifier Networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2877–2884.
 - [310] L. Kang, P. Ye, Y. Li, and D. Doermann, “Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks,” in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 2791–2795.
 - [311] D. Ghadiyaram and A. C. Bovik, “Massive Online Crowdsourced Study of Subjective and Objective Picture Quality,” *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 1–16, 2016.
 - [312] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, “On the Use of Deep Learning for Blind Image Quality Assessment,” *Signal, Image Video Process.*, vol. 12, no. 2, 2016.
 - [313] S. A. Amirshahi, M. Pedersen, and S. X. Yu, “Image Quality Assessment by Comparing CNN Features Between Images,” *Electron. Imaging*, vol. 12, pp. 1–19, 2017.
 - [314] P. Ye, J. Kumar, D. Doermann, and L. Kang, “Unsupervised feature learning framework for no-reference image quality assessment,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1098–1105.
 - [315] P. Ye, J. Kumar, L. Kang, and D. Doermann, “Real-time no-reference image quality assessment based on filter learning,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 987–994.
 - [316] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional Neural Networks for No-Reference Image Quality Assessment,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1733–1740.
 - [317] W. Lin and C. C. Jay Kuo, “Perceptual visual quality metrics: A survey,” *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.
 - [318] V. Pareto, *Cours d’Économie Politique*. Lausanne: L’Université de Lausanne, 1896.
 - [319] Wikipedia, “Airy Disk: Approximation using a Gaussian profile,” 2017. [Online]. Available: https://en.wikipedia.org/wiki/Airy_disk. [Accessed: 05-Feb-2017].

-
- [320] Apple Inc., “iPhone 6 - Technical Specifications,” 2017. [Online]. Available: https://support.apple.com/kb/sp705?locale=en_GB. [Accessed: 03-Dec-2016].
 - [321] Wikipedia, “Image sensor format,” 2017. [Online]. Available: https://en.wikipedia.org/wiki/Image_sensor_format. [Accessed: 01-Feb-2017].
 - [322] MathWorks, “Imnoise: Add noise to image,” 2017. [Online]. Available: <http://uk.mathworks.com/help/images/ref/imnoise.html>. [Accessed: 07-Apr-2018].
 - [323] Mathworks, “Imgaussfilt: 2-D Gaussian filtering of images,” 2017. [Online]. Available: <https://uk.mathworks.com/help/images/ref/imgaussfilt.html>. [Accessed: 07-Feb-2017].
 - [324] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing using MATLAB*, 1st ed. New Jersey, USA: Pearson Prentice Hall, 2004.
 - [325] S. Eddins, “Fourier transform visualization using windowing,” *MathWorks Blogs*, 2009. [Online]. Available: <https://blogs.mathworks.com/steve/2009/12/04/fourier-transform-visualization-using-windowing/>. [Accessed: 12-Feb-2018].
 - [326] *ISO/IEC 15948:2003 (E): Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification*. World Wide Web Consortium (WC3), 2003.
 - [327] E. W. Jin, B. W. Keelan, J. Chen, J. B. Phillips, and Y. Chen, “Softcopy quality ruler method: implementation and validation,” in *Proc. SPIE 7242, Image Quality and System Performance VI*, 2009, pp. 724206-724206–14.
 - [328] Eizo Nanao Corporation, *User’s Manual: ColorEdge CG245W Calibration Color LCD Monitor*. Eizo Nanao Corporation, 2010.
 - [329] *IEC/4WD 61966-2-1: Colour Measurement and Management in Multimedia Systems and Equipment - Part 2-1: Default RGB Colour Space - sRGB*. International Electrotechnical Commission (IEC), 1998.
 - [330] J. Schneider, “Snellen Near Vision Test Card,” 2002. [Online]. Available: <http://plantphys.info/sciencematters/snellenchart.pdf>. [Accessed: 03-Mar-2018].
 - [331] IEEE Standards Association, “IEEE Standards Downloads and Executable Files:

-
- IEEE 1858™,” 2016. [Online]. Available:
https://standards.ieee.org/content/dam/ieee-standards/standards/web/download/1858-2016_downloads.zip. [Accessed: 27-Oct-2018].
- [332] D. J. Baxter and A. Murray, “Calibration and adaptation of ISO visual noise for I3A’s Camera Phone Image Quality initiative,” in *Proc. SPIE 8293, Image Quality and System Performance IX*, 2012, no. May, pp. 829303-1-829303–14.
- [333] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “The SSIM Index for Image Quality Assessment.” [Online]. Available:
<https://ece.uwaterloo.ca/~z70wang/research/ssim/>. [Accessed: 11-Sep-2019].
- [334] T. T. Norton, D. A. Corliss, and J. E. Bailey, *The Psychophysical Measurement of Visual Function*. Butterworth-Heinemann, 2002.
- [335] S. Triantaphillidou, J. Smekjal, E. W. S. Fry, and C. H. Hung, “Studies on effect of megapixel sensor resolution on displayed image quality and relevant metrics,” in *Proc IS&T Electronic Imaging, Image Quality and System Performance XVII*, 2020.
- [336] Image Engineering, “iQ Automator Solution,” 2019. [Online]. Available:
<https://www.image-engineering.de/products/solutions/all-solutions/iq-automator>. [Accessed: 09-May-2019].
- [337] Imatest, “Sofica IQLaR Image Quality Lab Automation and Robotics,” 2019. [Online]. Available: <http://store.imatest.com/equipment/robotic-automation/sofica-iqlar-image-quality-laboratory-automation-and-robotics.html>. [Accessed: 09-May-2019].
- [338] B. Srinivasa Reddy and B. N. Chatterji, “An FFT-based Technique for Translation, Rotation, and Scale-Invariant Image Registration,” *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, 1996.
- [339] M. McGuire, “An image registration technique for recovering rotation, translation and scale parameters,” in *Technical Report 98-018, NEC Research Institute, Technical Report*, 1998, pp. 1–29.
- [340] G. Tzimropoulos, V. Argyriou, S. Zafeiriou, and T. Stathaki, “Robust FFT-based

-
- scale-invariant image registration with image gradients,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1899–1906, 2010.
- [341] O. van Zwanenberg, S. Triantaphillidou, R. Jenkin, and A. Psarrou, “Edge Detection Techniques for Quantifying Spatial Imaging System Performance and Image Quality,” in *Proc. IEEE New Trends in Image Restoration and Enhancement (NTIRE) workshop, in conjunction with Conference on Computer Vision and Pattern Recognition (CVPR) 2019*, 2019.
- [342] P. Kane, “Signal Detection Theory and Automotive Imaging,” in *Proc. IS&T Electronic Imaging, Autonomous Vehicles and Machines Conference*, 2019, pp. 027-1-027–7.
- [343] Vargas Aguilera C A, “LanczosFilter.m,” 2019. [Online]. Available: <https://uk.mathworks.com/matlabcentral/fileexchange/14041-lanczosfilter-m>. [Accessed: 19-Nov-2019].
- [344] S. Triantaphillidou and R. E. Jacobson, “Measurements of the modulation transfer function of image displays,” *J. Imaging Sci. Technol.*, vol. 48, no. 1, pp. 58–65, 2004.
- [345] E. W. S. Fry, S. Triantaphillidou, R. B. Jenkin, R. E. Jacobson, and J. R. Jarvis, “Noise Power Spectrum Scene-Dependency in Simulated Image Capture Systems,” in *Proc IS&T Electronic Imaging, Image Quality and System Performance XVII (submitted, under review)*, 2020.
- [346] M. Mody *et al.*, “Image Signal Processing for Front Camera based Automated Driver Assistance System,” in *IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2015, pp. 158–159.
- [347] M. Mody *et al.*, “High Performance Front Camera ADAS Applications on TI’s TDA3X Platform,” in *Proc. 22nd IEEE International Conference on High Performance Computing (HiPC)*, 2015, pp. 456–463.
- [348] Nvidia Corporation, *NVIDIA Tegra X1 Mobile Processor: Technical Reference Manual*. Nvidia Corporation, 2016.
- [349] R. Saussard, B. Bouzid, M. Vasiliu, and R. Reynaud, “Optimal performance

-
- prediction of ADAS algorithms on embedded parallel architectures,” in *Proc. IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security and 2015 IEEE 12th International Conference on Embedded Software and Systems*, 2015, pp. 213–218.
- [350] Texas Instruments, *Advanced Driver Assistance (ADAS) Solutions Guide*. Texas Instruments, 2015.
- [351] R. Jenkin and P. Kane, “Fundamental Imaging System Analysis for Autonomous Vehicles,” in *Proc. IS&T Electronic Imaging, Autonomous Vehicles and Machines Conference*, 2018, no. 17, pp. 105-1-105–10.
- [352] S. Dodge and L. Karam, “A study and comparison of human and deep learning recognition performance under visual distortions,” in *Proc. IEEE 26th International Conference on Computer Communications and Networks (ICCCN)*, 2017.
- [353] S. Diamond, V. Sitzmann, S. Boyd, G. Wetzstein, and F. Heide, “Dirty Pixels: Optimizing Image Classification Architectures for Raw Sensor Data,” *arXiv Prepr. arXiv1701.06487*, 2017.
- [354] R. B. Jenkin, “Comparison of Detectability Index and Contrast Detection Probability,” *J. Imaging Sci. Technol. (in print)*, 2020.
- [355] J. Y. Park, “Evaluation of changes in image appearance with changes in displayed image size,” PhD Thesis, Faculty of Media, Arts and Design, University of Westminster, UK, 2014.
- [356] *Recommendations on Uniform Color Spaces, Color Difference Equations, Psychometric Color Terms. Supplement No.2 to CIE Publication No.15 (E.-1.3.1) 1971/(TC-1.3)*. Commission Internationale de l’Eclairage (CIE), 1978.
- [357] *IEC 61966-4:2000: Multimedia systems and equipment — Colour measurement and management — Part 4: Equipment using liquid crystal display panels*. International Electrotechnical Commission (IEC), 2000.
